Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics

Teye, E., Amuah, C. L. Y., McGrath, T., & Elliott, C. (2019). Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *217*, 147-154. https://doi.org/10.1016/j.saa.2019.03.085

Published in:
Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

Document Version:
Peer reviewed version

Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics

*Ernest Teye[1&2], Charles. L. Y. Amuah[3], Terry McGrath[2] and Christopher Elliott[2]

[1]University of Cape Coast, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Institute for Global Food Security, Queen's University Belfast, Northern Ireland, U.K.

[3]University of Cape Coast, School of Physical Sciences, Department of Physics, Laser and Fibre Optics Centre, Cape Coast, Ghana

Tel :+233-243170302/+233-206969565

*Email: **teyernest@gmail.com** / **ernest.teye@ucc.edu.gh**

Abstract

Rice is the second most important food staple worldwide and the demand will continue to increase with the growth of the world population. As reports grow that frauds is prevalent in many supply chains there is the need for an effective and rapid technique for monitoring the authenticity and quality of rice. This study investigated the novel application of hand-held NIR spectrometry coupled to chemometric for the estimation of rice authenticity and quality in real time. A total of 520 Rice samples from different quality grades (high quality, mid quality and low quality) and different countries (Ghana, Thailand, and Vietnam) of origin were used. Among the pre-processing methods used multiplicative scatter correction (MSC) was found to be superior. Principal component analysis (PCA) was used to extract relevant information from the spectral data set and the results showed that rice samples of different categories could be clearly clustered under the first three PCs using the MSC preprocessing method. The performance of K-nearest neighbor (KNN) revealed that for authentication of rice quality grades, the classification rate gave 91.62% and 91.81% in training set and prediction set respectively while identification rate based on different country of origin was 90.84% and 90.64% in both training set and prediction set respectively. For the differentiation of local rice from the imported, KNN and SVM all had 100% in both the training set and prediction set. These gives very strong evidence that hand-held spectrometry coupled with MSC-PCA-KNN could successfully be used to

provide rapid and nondestructive classification of rice samples according to different quality grades, geographical origin and imported versus locally produced rice. This technique could enhance the work of quality control inspectors both from industry and regulatory perspectives for the rapid detection of rice integrity and fraud issues.

Keywords: Rice, Hand-held NIR spectroscopy, Chemometrics, Quality, Authenticity

1.0 INTRODUCTION

Rice (Oryza sativa L.) has increasingly become a hugely important staple food worldwide and its consumption has surged as a result of population growth, changing food preferences and urbanization [1]. Globally the top five leading net exporters of rice are Thailand, Vietnam, India, China, and Pakistan. Generally, Asian countries produce the largest quantities of rice worldwide while in Latin America, Middle East and African countries such as Ghana, Nigeria, Ivory Coast, has shown considerable increase in the importation and consumption of rice. Rice is consumed by more than 3.5 billion people (almost half of the world's population) and in 2015 West Africa countries imported more than 7.6 million tonnes of rice amounting to USD 4 billion [1]. Thus, Africa countries are net rice importers and given the increasing rice consumption trends in Africa this will continue to grow [2]. Due to the large diversity of rice in terms of types and country of origin and the increasing volumes being imported, it is extremely difficult to identify the quality and authenticity status at the ports and in the markets and has led to growing rumours that substandard product is flooding the African market places.

What is clear is that, some unscrupulous actors in the rice supply chain are selling sub-standard rice to consumers and very often these unscrupulous activities go undetected, because the existing analytical methods are cumbersome, time consuming, expensive and can involve substantial use of chemicals. For instance the DNA finger-printing method based on the genetic characteristic of rice has shown to be inappropriate for on-site application, it also involves labour intensive and skilled expertise [3, 4]. The procedure is also quite cumbersome and laboratory based and limits the number of sample that can be tested due to the cost involved. Other methods which are used are sensory examination by well-trained inspector [5, 6], enzyme linked immunosorbent assays, qualitative polymerase chain reaction, high performance liquid chromatography and gas chromatography [7, 8]. However all these methods has many

shortcomings, such as high cost, difficulty in use, and substantial time gap between sampling and results generation [9].

In a competitive market however, fast, accurate evaluation and classification of rice quality would lead to better decision making in terms of better quality control management. This would further give consumers a stronger assurance around product quality. NIR spectroscopy could provide this technology for such rapid examination. NIR spectroscopy is an excellent, rapid analytical method that has been used in several industries such as pharmaceutical, petrochemical, agricultural, and food processing. This method coupled with chemometrics has far-fetching advantages over the traditional analytical methods; it is rapid, nondestructive, reliable and involves no chemical usage [10, 11]. Also it requires little or no sample preparation and can be used in an on-line monitoring tool. Previous research has demonstrated that NIR spectroscopy has high potential for nondestructive measurement of qualitative and quantitative attributes in different agricultural and food products [12, 13].

For rice quality evaluation, NIR spectroscopy has been used for; discrimination of rice [9, 14], rice quality measurement [15-18] and classification of cultivars [19], authenticity detection [20], prediction of protein and amylase content [17, 21], detection of wax rice [22] and prediction of eating quality [23]. However, with the recent development of miniaturized or portable NIR spectrometers, there are no studies to date on the use of portable NIR devices for rice quality control. However a few studies have investigated the technology in terms of its use for food product quality measurement such as: identification of barley, chickpea and sorghum cultivar [24], determination of quality parameters in fruits; mango & pineapple [11, 25], determination of sucrose in sugar beet [26] and identification of coffee [27]. However, no studies until that reported herein are available for the use of hand-held NIR spectroscopy for rapid and nondestructive examination of rice quality and authenticity. Furthermore, there is no discussion on the use of different preprocessing techniques and multivariate classification methods for accurate determination of rice quality and authenticity. With the current incidences of rice fraud being reported worldwide such as in the media about ‘plastic rice’ and ‘toxic rice’ etc, this research could provide the urgently required reliable analytical technique for rapid detection of rice quality and authenticity. It would be particularly helpful in developing countries where the majority of these regions suffer from poor laboratory infrastructure and inadequate technical

know-how. The current research therefore seeks to develop a very reliable, nondestructive prediction model using a hand-held NIR spectrometer coupled with chemometrics for rapid and nondestructive evaluation of rice quality and authenticity. Also, there is the potential of this model to be imported into smart phone base for on-site applications.

2.0 MATERIALS AND METHODS

2.1 Sample preparation

In this study, rice samples were collected from local millers, and recognized retailers around seven regions in Ghana, while the other samples (imports from Thailand and Vietnam) were bought from recognised super-markets in Ghana. In all a total of 520 samples of rice grains were used in the study. For quality grades: 161 high quality, 250 mid quality and 109 low quality. For country of origin: 329 Ghana, 99 Thailand, 92 Vietnam, whiles the others were 191 imported samples and 329 local samples. These samples were bagged in sealable polythyene bags and were grouped into three quality categories namely; high quality, mid quality and low quality. These quality groupings were based on the cost price, and confirmed by sensory analysis using a skilled panelist to exam the physical appearance, cleanliness, colour, and aroma on a six (1-6) corresponding level as done by other authors [28]. All samples were (three replication each) were sent to the laboratory for further analysis.

2.2 Sample spectral acquisition

The spectrum of each rice sample was collected using a hand-held spectrometer (SCIO^{TM}) with spectral range between 740 nm and 1070 nm in a 1-nm resolution for spectral data recording. Samples (50 g) were collected into glass containers and scanned three times after rotating the cup. The whole process was carried out at ambient temperature. The set-up of the scanning procedure is shown in the graphical abstract.

2.3 Software device

Spectral data recordings stored in a cloud-based dataset with their corresponding reference value for time of scanning were downloaded using a research license acquired from SCIO lab and imported to Matlab version 9.5.0 (Mathworks Inc., USA) with windows 10 Basic for data processing (All preprocessing treatments and multivariate algorithms).

2.4 Spectral preprocessing techniques

The raw spectra profile from the rice samples are shown in Fig. 1.0 (A) and the mean spectra of the various categories of the rice samples of interest are detailed in Fig. 1.0 (B to C). This outlines the three categories of rice samples used in the study. The average of the triplicate scans (whole spectral data set) were preprocessed before further analysis. The activity of preprocessing the spectra data is an integral part of modelling to eliminate background information and noise from the useful properties of the scanned samples [29, 30]. In this study, four spectra preprocessing techniques were applied comparatively, namely: de-trending (DT), mean centering (MC), multiplicative scatter correction (MSC), and standard normal variant (SNV) were employed as the specral models developed using unprocessed did not yield good results.These few preprocessing techniques were selected after initial background studies. These preprocessing techniques were undertaken so that the results from the model were based on the chemical fingerprint from the spectral information acquired.

2.5 Theory of chemometric techniques

2.5.1 De-trend (DT)

DT was trailled because it is frequently applied to spectra to remove baseline drift and curvilinearity for densely packed solids [31]. In detrending, a trend line is made from spectral data set by least squares fitting and then the trend line is subtracted from the original spectrum to correct for any defect.

2.5.2 Mean centering (MC)

MC performs calculation by averaging the spectrum of the data set and subtracting the average from each spectrum [31].

2.5.3 Multiplicative scatter correction (MSC)

MSC is a well known technique used since 1983 for the removal of undesirable scatter effect from spectra data matrix before modelling [29, 32]. It is also used to compensate the effect of non-uniform scattering induced by diverse particle size, uneven distribution and other effects in the spectral data and it works by linearizing each spectrum to an “ideal”spectrum, which

corresponds to the average spectrum of the calibration set and the average spectrum is fitted through the method of least squares [33].

2.5.4 Standard normal variant (SNV)

SNV is similar to MSC, it is a transformational spectral treatment method normally applied to remove the multiplicative interferences of scatter, particle size and the change of light disance [34]. In SNV each individual spectrum is normalized to zero mean and unit variance [13, 34].

2.5.5 Principal component analysis (PCA)

PCA is an unsupervised pattern recognition method that; employs the technique of extracting information from correlation matrices to visualize data trends in a dimensional scatter plot. The unsupervised classification terminology means that; the samples are classified with no prior knowledge, except the sensor signals. PCA expresses information contained in a dataset by principal components or reduces the dimension of the data matrix and compresses the information into interpretable variables called principal components (PCs) which are orthogonal [35, 36]. PCA has been used to observe possible groupings in several sensor data set. The best performing PCs normally show the most important information. Therefore, similar samples are grouped closer to each other and vice versa. Graphical profile of PCA results normally provides an initial output for determining the possible differences and similarities in a data set. PCA can be used to identify combinations of variables that have the largest contribution to variations in the data set, as these variables are retained in the first two or three PCs [36].

2.6 Data partition

The spectral data-set (520 samples) was downloaded and preprocessed with suitable techniques. Furthermore, these data were divided into two subsets called: calibration set (347 samples); was used to develop the model and prediction set (173 samples); was used for evaluating the actual predictive ability of the constructed models. The members in each set were selected in order to come to a 2/1 division of calibration set/prediction set. To avoid bias, members of the subset were selected as follows: for every three samples, two spectra were randomly selected as the calibration set while the remaining sample was the prediction set.

2.7 Multivariate classification models

The advancement in computers and electronics have contributed to making multivariate calibration a very powerful tool for processing NIR spectral data as it overcomes the difficulty of multi-collinearity and gives scientific statistical inferences for meaningful conclusions to experimental results [37, 38]. However, choosing the best method can be quite a cumbersome process as many are in existence. In this study; linear and non-linear algorithm were employed such as; K-nearest neighbor (KNN) and Support vector machine (SVM) respectively.

2.7.1 K-nearest neighbour (KNN)

KNN is a linear and non-parametric tool [39] which works based on a distance function that measures the difference or similarity between two stances [40]. KNN classifies an unknown sample of the validation set according to the class which belongs to the majority of its K nearest neighbour in the training set. In KNN, parameter K has a great impact on the classification model; hence the choice of K is normally optimized by calculating the prediction potential with several K values, preferably an odd number of small K values (3 or 5). Also to improve the performance of KNN model, there is the need to simultaneously optimize PCs and parameter K to derive a very good model. Furthermore, KNN technique presents some advantages such as it is: free from statistical assumption, mathematically simple but achieving classification results as good as other more complex pattern recognition method, and its effectiveness does not depend on the space distribution of the classes. However, it is known that KNN cannot work well if large differences are present in the number of samples in each class [41]. This therefore makes KNN tool a suitable technique for modeling similar class groupings. Other authors used KNN for modeling identification of tea grade groups [42] and discrimination of roasted tea [43].

2.7.2 Support vector machine (SVM)

SVM is supervised statistical learning theory applicable for both classification and regression problems or ranking functions [44]. It has shown good performance for classifying high-dimensional data when a limited number of training samples are available [45]. SVM processes sensor data by obtaining the optimal boundary of two groups in a vector space independent on the probabilistic arrangements of vectors in the training set. When the linear boundary in the low dimension input space is not enough to separate the two classes, SVM can create a hyperplane that allows linear separation in the higher dimension feature space. However, if the classes are separated by non-linear boundary, then the kernel function is used to find the boundary by

mapping the non-separable data into a higher dimensional space and causes the classes to become linearly separated. The strength of SVM over the others is that; it can achieve higher generalization by maximizing the margin and it can support an efficient learning of non-linear functions using the kernel trick [44]. Among the three kernel functions (sigmoid kernel, polynomial kernel and Gaussian kernel), Gaussian kernel function is mostly employed because of its simplicity and speed during its computation [41]. SVM is likely to have a better generalization, thus can accurately classify testing data points. For more information refer [44].

2.8 Model evaluation procedure

In this study, all the multivariate classification models were tested using the performance measure as the co-efficient of determination in a validation set for discrimination rate (%)

3.0 RESULTS AND DISCUSSION

3.1 Spectral presentation

All the rice samples used in this study were scanned three times from different angles after rotating the sampling cup 45^0 clockwise. The fingerprint from the spectra was used to create the statistical models. Figure 1 shows A= raw spectra, B= average spectra for quality categories, C= average spectra for country of origin and D= imported versus local rice samples. As seen from this figure, the spectra share very similar absorbance patterns in the 740 – 1070 nm range. It appears they could not be easily distinguished by interpretation of the raw data. However, these spectra did contain very useful yet unexploded information, therefore the need to pretreat the spectra data with suitable pretreatment techniques was apparent. The selection of the most suitable pretreatment techniques can be quite cumbersome, and this was mainly performed using a trial and error approach involving several selections [29, 46]. Finally, three preprocessing techniques were selected and were compared with unprocessed (raw) spectral profile.

Figure 1.0 NIR spectra of rice samples; (A) = raw, (B) = mean of different quality categories, (C) = country of origin and (D) = imported versus local

3.2 Principal component analysis (PCA)

The PCA results from the three attributes studied; quality grades, countries of origin and imported versus locally grown revealed that the topmost first three PCs gave clear separations in all the attributes studied with MSC providing the best preprocessing results in PCA among the

three others methods (Raw, MC, SNV). For MSC-PCA method, all the samples clustered well along the first three PCs plane.

3.3 Classification by quality grades

The score plots obtained after MSC+PCA with all the spectral fingerprints based on the quality grades are shown in Figure 2. The first two principal components (PCs); PC1 and PC2 can explain 86.78% and 7.63% of the variance respectively giving a total accumulation contribution of 94.41% total variance for the 520 samples used in this study. A neat clustering of the quality grades were observed for MSC+PCA while a faint clustering was observed for MC, De-trending and SNV. This phenomenon could be explained as MSC has the ability to correct scattering effect. KNN and SVM were applied separately after PCA to perform the identification between the three quality grades. The results of the identification rate are summarized in Table 1. From this table, it can be seen that the derived model; MSC+PCA+KNN gave an optimal identification rate above 90% in the training set and prediction set as compared to the others.

Figure 2.0 PCA score plot of rice sample of different quality preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC

Table 1.0 Performance of multivariate classification models based on quality grades

3.4 Classification by country of origin

After the identification of the quality grades, the possibility to discriminate between rice from different countries was investigated. This is particularly important because some consumers prefer rice from other countries over others. The variance in the data was mainly due to the country of origin (Ghana, Thailand, and Vietnam). For the PCA score plot shown in Figure 3, the best in-term of neat clustering was MSC-PCA where PC1, PC2 and PC3 can explain 80.97%, 14.15% and 2.09% of the variance respectively. With a total accumulation contribution of 97.22% variance for the 520 samples used. KNN and SVM were applied on all the samples to perform the discrimination of countries of origin. The results of the discrimination are summarized in Table 2. From this table it can be seen that model MSC+PCA+KNN gave the best identification rate of 90.84% and 90.64% in both the training set and prediction set as compare to the other models.

Figure 3.0 PCA score plot of rice sample from different origin preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC

Table 2.0 Performance of multivariate classification models based on country of origin

3.5 Classification of imported versus local

After the classification of quality grades and country of origin, the ability to differentiate between imported rice and local produce was investigated. This is particularly useful with the current incidences of imported rice fraud often reported in the media. Also the total ban on rice importation by some African countries and the encouragement of local production to reduce imports calls for easy and inexpensive testing techniques. Furthermore, cheap imports are competing with high quality local brands. Hence the potential of food fraudsters operating to gain economic advantage is large. For imported versus local (as seen from Figure 4) PC1, PC2 and PC3 gave 98.89%, 0.84% and 0.11% of the variance respectively, giving a total accumulation contribution of 99.84% variance for the 520 samples used. Again MSC+PCA gave a visible cluster trend in the PCA score plot of imported rice versus local rice. The derived model; MSC+PCA+KNN and MSC+PCA+SVM for differentiating imported from local gave 100% differentiation rate in both the training set and prediction set at 2 PCs respectively (as seen from Table 3) compared to the others. The results proved that KNN and SVM could be an excellent model for the classification of rice samples from different origins.

Figure 4.0 PCA score plot of rice sample local & imported preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC

Table 3.0 Performance of multivariate classification models based on imported versus local

4.0 General discussion

The NIR scanning process on the rice produced spectra that showed multiple bands and a few peaks as seen from the spectral profile presented. These bands are made up of overtones and combinations of fundamental vibrations which correspond to organic properties that provide unique characteristics or a fingerprint of the rice samples used. To the naked eye they appear very similar, though there exists many differences with useful and non-useful information. To extract the useful and unique information, chemometric techniques were applied. Firstly several

(MC, SNV, DT, MSC) preprocessing treatments were performed and compared prior to PCA. PCA as an unsupervised mathematical technique was used for extracting information from correlation matrices [47]. This gave visible cluster trends to show the strength of each preprocessing technique as shown in Figure 2, 3 & 4. MSC was found to give neat cluster trends and hence superior to the others applied in this study. This could be explained in that, MSC has the power to remove undesirable scatter effect of spectra from a data matrix before modelling [29] and this has proved highly useful in this study. Also, the scatter effect is known to influence modelling as it contains unwanted information for prediction. The groupings found in the PCA cluster plot could then be explained further by the chemical properties in each as a result of the differences in quality grade, geographical origin, pre-harvest and postharvest practices [31]. The contribution of the best performing three PCs were 98.67%, 97.22% and 99.84% for quality grade, country of origin and imported versus local respectively. However, PCA is not a classification tool but only functions by reducing dimensionality while preserving as much variance in a high dimensional space. Thus making it possible to extract useful information from high-dimensional data since the spectral data forms the array of correlated variables which contains overlapped information [47]. Hence, the preprocessed data (MSC-PCA data; which provided useful unique information) were further developed by a supervised pattern recognition technique known as KNN and SVM. The classification rates for the KNN models were above 90% in both the training set and prediction set respectively for all the types of rice studied. This could be explained by the fact that, KNN; as a simple linear and non-parametric tool classified the groups according to the classes which belongs to the majority of its K nearest neighbor. In this study PCs were also used as an input data in KNN and this enhanced the efficiencies of the KNN models. This was because factors (PCs & K) were simultaneously optimized by cross-validation method [43] as this act is useful for deriving a very good KNN model [31]. Also the number of PCs used in the KNN models suggests that it is simpler and normally, lower values of parameter K (3 or 5) and PCs are preferred as higher number of PCs included in training a model brings out too much redundant information which inescapably influences the robustness of the model [38, 48]. KNN is also known to perform equally to (or even better than) other more complex pattern recognition techniques but its effectiveness does not depend on the space distribution of the classes [41]. The performance of KNN model in this study was similar to the findings by other authors; classification of tea quality grades [42] discrimination of roasted tea

[43]. However, KNN cannot work well if large differences are present in the spectral dataset as it is known to be a 'lazy' training algorithm [41]. SVM was also employed in this study. From the results, SVM model also showed comparatively good results as KNN with 100% classification rate in both the training set and prediction set for only classifying imported versus local as shown in Table 3. This optimal result could be explained in that, SVM model embodies structural risk minimization principle where upper boundary is reduced on the expected risk and also possess the power of a better generalization, thus classifying accurately in the prediction set [43, 44, 49]. Comparing the two models (KNN and SVM) it must be emphasized that since KNN being a lazy learning method and the class probability estimation is based on a simple voting of a "good value" for k, as well as stores all the training data, the prediction stage is often slow hence require high memory and quite expensive [40, 50, 51]. However in this study it proved very useful for all the classification problems.

5.0 CONCLUSION

This study has revealed, for the first time, that hand-held NIR spectrometer coupled with the appropriate chemometrics could be used for rapid and nondestructive detection of rice authenticity and quality. The systematic selection of different preprocessing methods (MC, DT, SNV, and MSC) with PCA and modeling with KNN and SVM multivariate calibration model showed that MSC+PCA plus KNN showed superiority in this study with more than 90% classification rate for all categories of rice samples studied. Generally, it could be concluded that hand-held spectrometer together with appropriate multivariate calibration model could be exploited for quick and nondestructive detection of rice quality and authenticity. Furthermore, there is a potential for this model to be imported into smart phone for effective quality control measurements in the rice industry and by regulators as compared with the time-consuming wet chemistry analytical methods. Further studies are required for the prediction of other quality attributes in rice.

Acknowledgement

The authors highly appreciate the financial support provided by Agilent foundation (sub-contract No. 4368). Proof reading assistance provided by Mrs. Winifred Akpene Teye is also acknowledged. The kind contribution given by University of Cape Coast is fully appreciated.

References

1. Fiamohe, R., et al., *How Can West African Rice Compete in Urban Markets? A Demand Perspective for Policymakers.* EuroChoices, 2018.
2. Kormawa, P., et al. *Global Rice Trade: Dynamics, Policy Conflicts and Strategies in Africa*. in *Conference on International Agricultural Research for Development. Herausgegeben von Africa Rice Center (WARDA). Cotonou*. 2005.
3. Shanthi, P., et al., *DNA finger printing of salt tolerant and susceptible genotypes using microsatellite markers in rice (Oryza sativa L.).* Int J Plant Breed Genet, 2012. **6**: p. 206-16.
4. Jeung, J., et al., *Fingerprinting temperate japonica and tropical indica rice genotypes by comparative analysis of DNA markers.* Euphytica, 2005. **146**(3): p. 239-251.
5. Kwak, H.S., et al., *Identification of key sensory attributes for consumer acceptance and instrumental quality of aseptic‐packaged cooked rice.* International Journal of Food Science & Technology, 2015. **50**(3): p. 691-699.
6. Champagne, E.T., et al., *Important sensory properties differentiating premium rice varieties.* Rice, 2010. **3**(4): p. 270-281.
7. Wu, J., et al., *Metabolite profiles of rice cultivars containing bacterial blight-resistant genes are distinctive from susceptible rice.* Acta Biochim Biophys Sin, 2012. **44**(8): p. 650-659.
8. Kemnitz, D., S. Kolb, and R. Conrad, *Phenotypic characterization of Rice Cluster III archaea without prior isolation by applying quantitative polymerase chain reaction to an enrichment culture.* Environmental Microbiology, 2005. **7**(4): p. 553-565.
9. Long, Z., et al., *Discrimination of transgenic rice based on near infrared reflectance spectroscopy and partial least squares regression discriminant analysis.* Rice science, 2015. **22**(5): p. 245-249.
10. Cen, H., Y. He, and M. Huang, *Combination and comparison of multivariate analysis for the identification of orange varieties using visible and near infrared reflectance spectroscopy.* European Food Research and Technology, 2007. **225**(5-6): p. 699-705.
11. Marques, E.J.N., et al., *Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer.* Food chemistry, 2016. **197**: p. 1207-1214.
12. Huang, X., et al., *Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue.* Analytical Methods, 2014. **6**(14): p. 5008-5015.
13. Nicolai, B.M., et al., *Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review.* Postharvest biology and technology, 2007. **46**(2): p. 99-118.
14. Chen, K. and M. Huang, *Prediction of milled rice grades using Fourier transform near-infrared spectroscopy and artificial neural networks.* Journal of cereal science, 2010. **52**(2): p. 221-226.
15. Chang, R., et al. *Research of rice-quality based on computer vision and near infrared spectroscopy*. in *International Conference on Computer and Computing Technologies in Agriculture*. 2009. Springer.
16. Caporaso, N., M.B. Whitworth, and I.D. Fisk, *Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains.* Applied Spectroscopy Reviews, 2018: p. 1-21.
17. Zhang, B., et al., *Prediction of the amino acid composition in brown rice using different sample status by near-infrared reflectance spectroscopy.* Food Chemistry, 2011. **127**(1): p. 275-281.
18. Natsuga, M. and S. Kawamura, *Visible and near-infrared reflectance spectroscopy for determining physicochemical properties of rice.* Transactions of the ASABE, 2006. **49**(4): p. 1069-1076.
19. Kong, W., et al., *Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariate data analysis.* sensors, 2013. **13**(7): p. 8916-8927.
20. Chen, H., C. Tan, and Z. Lin, *Authenticity Detection of Black Rice by Near-Infrared Spectroscopy and Support Vector Data Description.* International journal of analytical chemistry, 2018. **2018**.

21. Xie, L., et al., *Optimisation of near-infrared reflectance model in measuring protein and amylose content of rice flour.* Food chemistry, 2014. **142**: p. 92-100.
22. Li, B., et al., *Detection of waxed rice using Visiblenear infrared hyperspectral imaging.* J. Food Nutr. Res, 2016. **4**: p. 267-275.
23. Siriphollakul, P., et al., *Eating quality evaluation of Khao Dawk Mali 105 rice using near-infrared spectroscopy.* LWT-Food Science and Technology, 2017. **79**: p. 70-77.
24. Kosmowski, F. and T. Worku, *Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia.* PloS one, 2018. **13**(3): p. e0193620.
25. Chia, K.S., H.A. Rahim, and R.A. Rahim, *Prediction of soluble solids content of pineapple via non-invasive low cost visible and shortwave near infrared spectroscopy and artificial neural network.* Biosystems Engineering, 2012. **113**(2): p. 158-165.
26. Pan, L., et al., *Determination of sucrose content in sugar beet by portable visible and near-infrared spectroscopy.* Food chemistry, 2015. **167**: p. 264-271.
27. Correia, R.M., et al., *Portable near infrared spectroscopy applied to quality control of Brazilian coffee.* Talanta, 2018. **176**: p. 59-68.
28. Lu, L., et al., *Determination of rice sensory quality with similarity analysis-artificial neural network method in electronic tongue system.* RSC Advances, 2015. **5**(59): p. 47900-47908.
29. Coronel-Reyes, J., et al., *Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques.* Computers and Electronics in Agriculture, 2018. **145**: p. 1-10.
30. Blanco, M. and I. Villarroya, *NIR spectroscopy: a rapid-response analytical tool.* TrAC Trends in Analytical Chemistry, 2002. **21**(4): p. 240-250.
31. Teye, E., et al., *Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2013. **114**: p. 183-189.
32. Martens, H., S. Jensen, and P. Geladi. *Multivariate linearity transformation for near-infrared reflectance spectrometry*. in *Proceedings of the Nordic symposium on applied statistics*. 1983. Stokkand Forlag Publishers Stavanger, Norway.
33. Wang, H., et al., *Fruit quality evaluation using spectroscopy technology: a review.* Sensors, 2015. **15**(5): p. 11889-11927.
34. Barnes, R., M.S. Dhanoa, and S.J. Lister, *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* Applied spectroscopy, 1989. **43**(5): p. 772-777.
35. Hong, X. and J. Wang, *Detection of adulteration in cherry tomato juices based on electronic nose and tongue: Comparison of different data fusion approaches.* Journal of Food Engineering, 2014. **126**: p. 89-97.
36. Lavine, B.K. and N. Mirjankar, *Clustering and classification of analytical data.* Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation, 2006.
37. Roggo, Y., et al., *A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies.* Journal of pharmaceutical and biomedical analysis, 2007. **44**(3): p. 683-700.
38. Teye, E., et al., *Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis.* Food chemistry, 2015. **176**: p. 403-410.
39. Duda, R.O., P.E. Hart, and D.G. Stork, *Pattern classification*. 2012: John Wiley & Sons.
40. Jiang, L., et al. *Survey of improving k-nearest-neighbor for classification*. in *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*. 2007. IEEE.
41. Berrueta, L.A., R.M. Alonso-Salces, and K. Héberger, *Supervised pattern recognition in food analysis.* Journal of chromatography A, 2007. **1158**(1-2): p. 196-214.

42. Chen, Q., J. Zhao, and S. Vittayapadung, *Identification of the green tea grade level using electronic tongue and pattern recognition.* Food Research International, 2008. **41**(5): p. 500-504.
43. Chen, Q., J. Zhao, and H. Lin, *Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition.* Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2009. **72**(4): p. 845-850.
44. Yu, H. and S. Kim, *SVM tutorial—classification, regression and ranking*, in *Handbook of Natural computing*. 2012, Springer. p. 479-506.
45. Tarabalka, Y., et al., *SVM and MRF-based method for accurate classification of hyperspectral images.* IEEE Geoscience and Remote Sensing Letters, 2010. **7**(4): p. 736-740.
46. Rinnan, Å., F. van den Berg, and S.B. Engelsen, *Review of the most common pre-processing techniques for near-infrared spectra.* TrAC Trends in Analytical Chemistry, 2009. **28**(10): p. 1201-1222.
47. Mehdizadeh, S.A., et al., *An intelligent system for egg quality classification based on visible-infrared transmittance spectroscopy.* Information Processing in Agriculture, 2014. **1**(2): p. 105-114.
48. Chen, Q., et al., *Nondestructive identification of tea (Camellia sinensis L.) varieties using FT-NIR spectroscopy and pattern recognition.* Czech J Food Sci, 2008. **26**: p. 360-367.
49. Lin, H., et al., *Determination of free amino acid content in Radix Pseudostellariae using near infrared (NIR) spectroscopy and different multivariate calibrations.* Journal of pharmaceutical and biomedical analysis, 2009. **50**(5): p. 803-808.
50. Thirumuruganathan, S., *A detailed introduction to K-nearest neighbor (KNN) algorithm.* Retrieved on July, 2010. **21**: p. 2015.
51. Guo, G., et al. *KNN model-based approach in classification*. in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. 2003. Springer.

Table 1.0 Performance of multivariate classification models based on quality grades

Model		Number of principal components	Correct classification rate (%)	
			Training set (347)	Prediction set (173)
KNN	MC	7	69.59	76.02
	MSC	7	91.62	91.81
	DT	6	55.36	56.72
	SNV	5	77.39	78.95
SVM	MC	6	64.33	64.33
	MSC	5	87.52	86.55
	DT	5	48.34	47.37
	SNV	7	64.33	63.74

Table 2.0 Performance of multivariate classification models based on country of origin

Model		Number of principal components	Correct classification rate (%)	
			Training set (347)	Prediction set (173)
KNN	MC	7	75.63	72.51
	MSC	6	90.84	90.64
	DT	7	52.05	56.73
	SNV	5	86.35	87.36
SVM	MC	7	79.14	78.36
	MSC	6	82.46	83.63
	DT	5	64.91	64.91
	SNV	7	79.92	80.70

Table 3.0 Performance of multivariate classification models based on imported versus local

Model		Number of principal components	Correct classification rate (%)	
			Training set (347)	Prediction set (173)
KNN	MC	5	90.84	91.81
	MSC	2	100.00	100.00
	DT	5	64.72	66.08
	SNV	5	96.30	96.49
SVM	MC	5	95.32	95.32
	MSC	2	100.00	100.00
	DT	5	62.77	60.23
	SNV	7	95.91	97.08

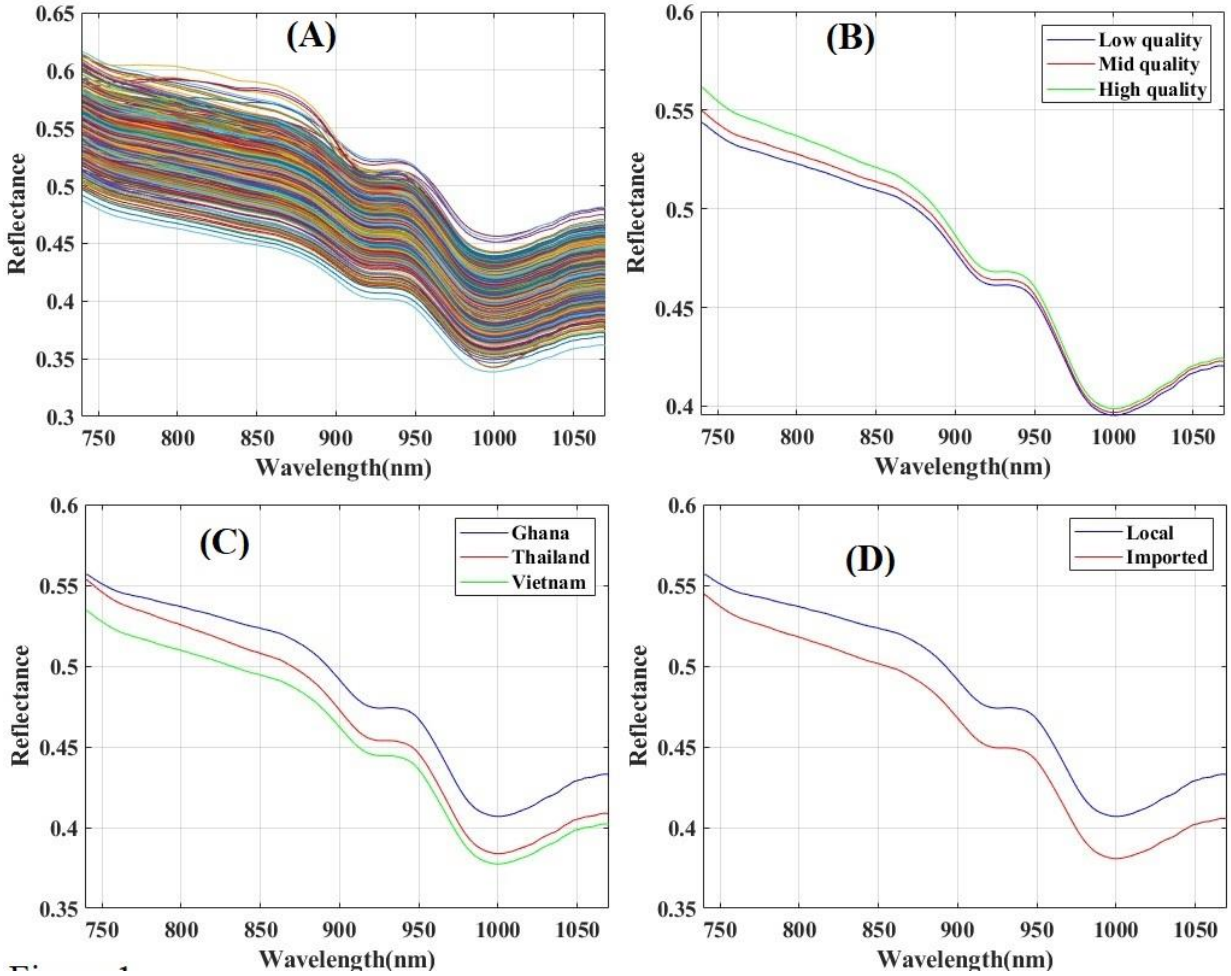

Figure 1. NIR spectra of rice samples; (A) = raw, (B) = mean of different quality categories, (C) = country of origin and (D) = imported versus local

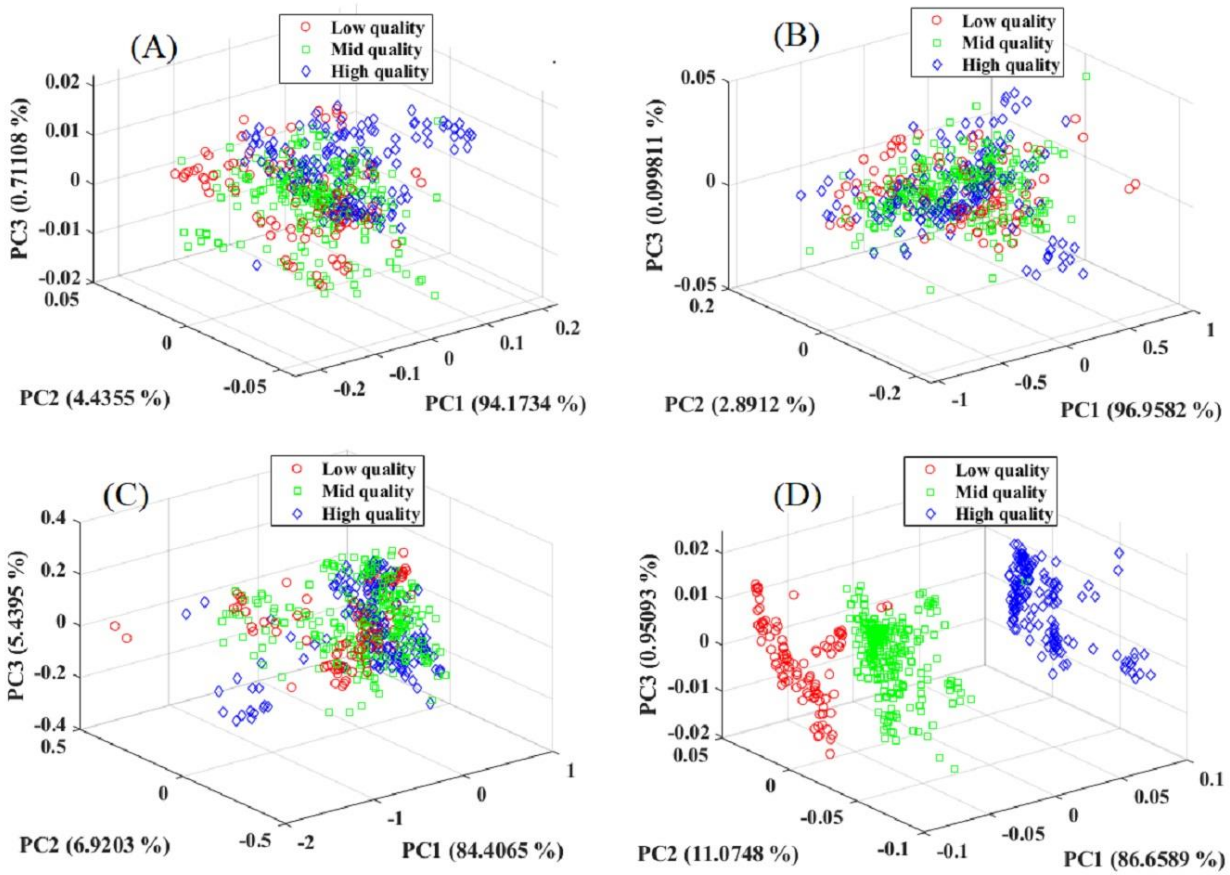

Figure 2. PCA score plot of rice sample of different quality preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC.

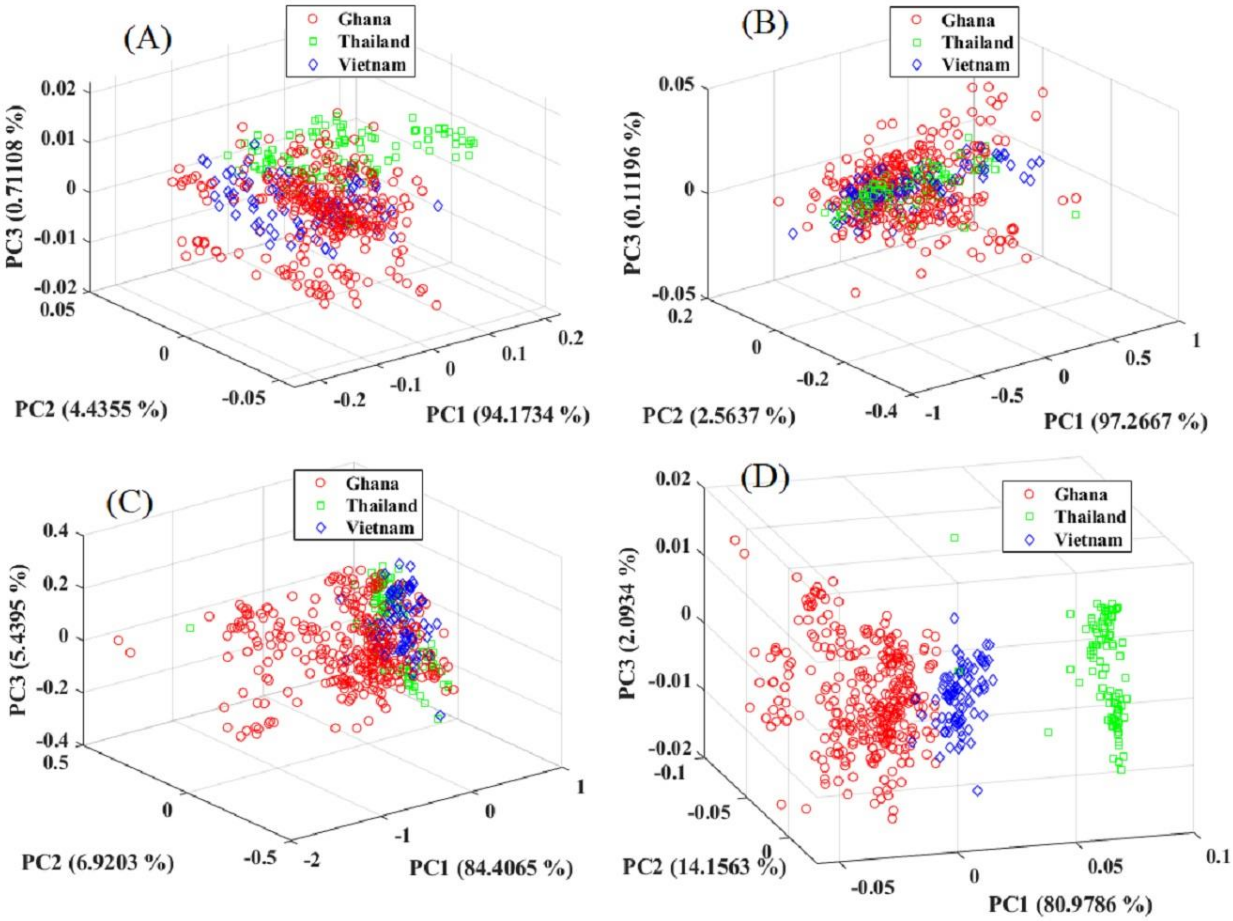

Figure 3. PCA score plot of rice sample from different origin preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC.

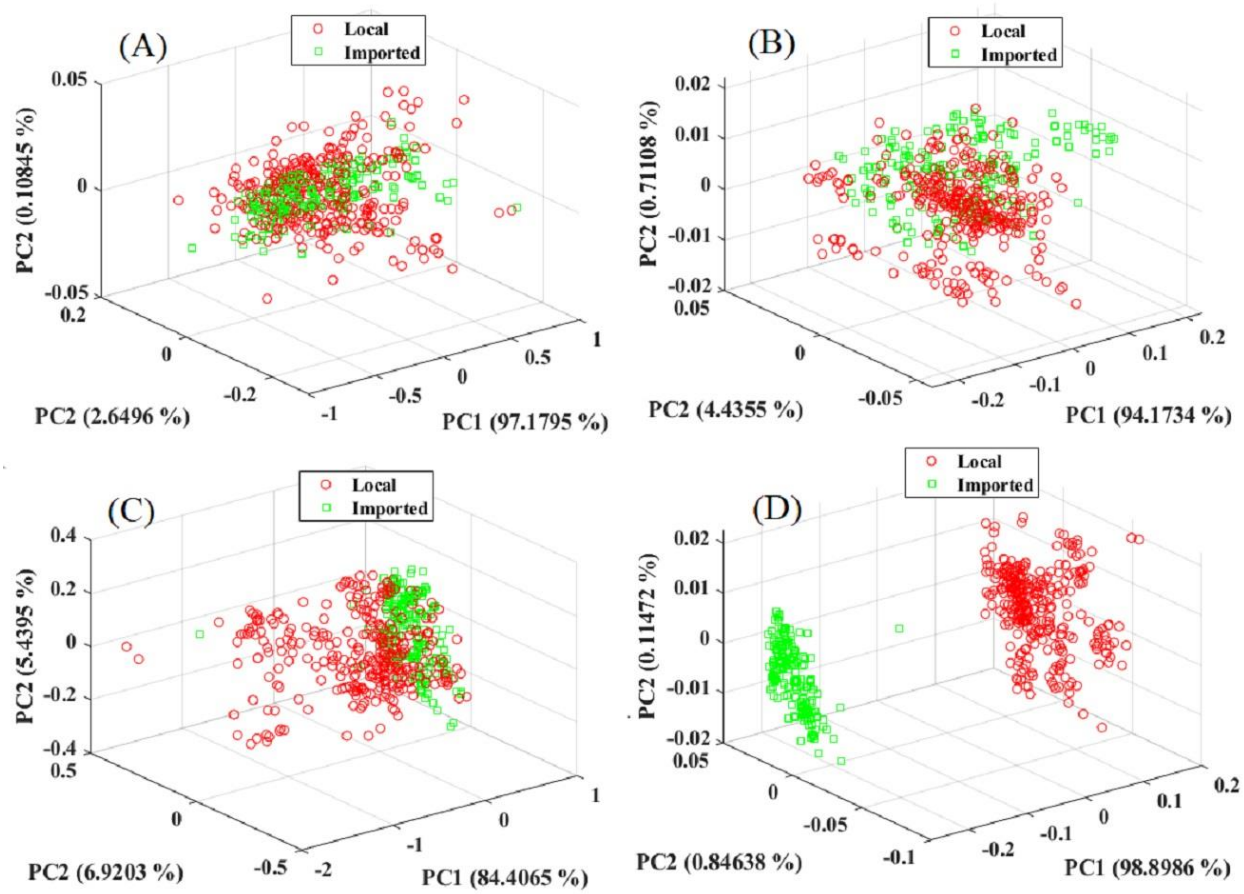

Figure 4. PCA score plot of rice sample local & imported preprocessed (A) MC, (B) DT, (C) SNV and (D) MSC.