Food Physics 2 (2025) 100059

Contents lists available at ScienceDirect

Food Physics

journal homepage: www.keaipublishing.com/en/journals/food-physics/

Rice seed integrity evaluation: Developing a rapid onsite system to check seed fraud using a portable NIR spectroscopic device coupled with smartphone technology

Ernest Teye[a,g,*], Charles Lloyd Yeboah Amuah[b], Vida Gyimah Boadu[c], Kwadwo Anokye Dompreh[b], Maxwell Darko Asante[d], Francis Padi Lamptey[e], Stephen Narh[f], Daniel Dzorkpe Gamenyah[d], George Oduro Nkansah[g], Selorm Akaba[h]

[a] *Department of Agricultural Engineering, School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, Ghana*
[b] *Department of Physics, Laser and Fibre Optics Centre, School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, Ghana*
[c] *Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Department of Hospitality and Tourism Education, Kumasi, Ghana*
[d] *CSIR-Crop Research Institute, Fumesua, Kumasi, Ghana*
[e] *Cape Coast Technical University, School of Applied Sciences and Technology, Department of Food Science and Postharvest Technology, Cape Coast, Ghana*
[f] *Soil and Irrigation Research Centre, School of Agriculture, University of Ghana, Legon, Ghana*
[g] *African Agribusiness Consortium, Jospong Group of Companies, East Legon, Accra, Ghana*
[h] *Department of Agricultural Economics and Extension, School of Agriculture, University of Cape Coast, Cape Coast, Ghana*

ARTICLE INFO

Keywords:
Seed fraud
NIR spectroscopy
Rice seed authentication
Chemometric analysis
Varietal identification

ABSTRACT

Rice seed integrity is critical in ensuring high yield and grain quality; however, seed fraud, particularly the misrepresentation of rice paddy (unhusked rice grain) as rice seed, is a growing concern that threatens sustainability efforts. This study investigates using a portable NIR spectroscopic device, combined with chemometric analysis, for rapid onsite identification of rice seed and paddy varieties for real-time verification of seed authenticity. A total of 280 rice samples, representing four varieties (Agra, Amankwatia, Legon 1, and Jasmine 85) across two categories (seeds and paddy), were analyzed. After applying various pre-processing techniques and principal component analysis (PCA), linear discriminant functions 1 and 2 successfully revealed distinct clustering patterns for both the varieties and categories (rice seed and paddy). Among the classification algorithms used, Random Forest (RF) achieved 100 % accuracy for rice seed identification and 97.38 % for paddy identification in the test sets. Support Vector Machine (SVM) demonstrated 98.15 % accuracy in distinguishing between rice seed and paddy for detecting seed fraud. These results suggest that a portable NIR device can reliably perform varietal identification and seed authenticity checks within the agricultural value chain. This technology has significant potential for use by seed inspectors, farmers, and regulatory officers, offering a non-destructive, real-time solution for the rice industry.

* Corresponding author at: Department of Agricultural Engineering, School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, Cape Coast, Ghana

E-mail addresses: eteye@aacint.org, ernest.teye@ucc.edu.gh (E. Teye).

https://doi.org/10.1016/j.foodp.2025.100059
Received 14 May 2025; Received in revised form 23 June 2025; Accepted 10 July 2025
Available online 11 July 2025

1. Introduction

Rice (*Oryza sativa* L.) production and consumption have been steadily increasing in Sub-Saharan Africa due to factors such as population growth, urbanization, convenience of preparation, and changing consumer preferences [17]. However, rice production in the region faces numerous challenges, resulting in lower yields compared to other regions [38,4,49]. One issue that has received limited attention is the integrity of rice seeds, which has opened the door for fraudsters to sell inferior seeds, a problem that often goes undetected, particularly at the farm gate [16,44,5]. Seed fraud is an increasingly concerning issue across Africa [28], where agriculture plays a central role in many economies and is vital to food security. Misrepresenting rice paddy as rice seed directly affects crop yields and threatens farmers' livelihoods [44]. Unlike genuine rice seeds that undergo rigorous quality control to ensure it meets standards, such as high germination potential and genetic purity, paddy (fake seeds) lack the necessary treatment and selection processes, such as drying, cleaning, and sorting, that are essential for producing viable seeds for planting.

Genuine rice seeds, specially bred and selected for planting, are often mixed or replaced with rice paddy, a cheaper and less viable alternative not intended for planting. This substitution results in lower germination rates, less disease resistance, and reduced productivity. Standard methods for distinguishing between rice seed and paddy, as well as assessing seed quality, such as germination tests [19,35,50], purity tests [1,26,33,56], DNA fingerprinting [20,42,43], enzyme analysis [25,27], and sensory evaluations [11,23,37] are time-consuming, require specialized expertise, are often destructive, and depend on laboratory facilities that may be inaccessible to farmers in rural areas [46]. Additionally, advanced methods, as described in previous publications [45], also present several limitations.

There is a need for fast, accurate, and onsite detection methods to address these challenges and contribute to food security by ensuring the integrity of rice seeds. Such methods would provide farmers and agricultural extension agents with a reliable tool to ensure seed integrity. Portable Near-Infrared (NIR) spectroscopy has emerged as a promising, cost-effective solution [14,51,52,6], particularly in regions where access to advanced laboratory infrastructure is limited, and the need for immediate, reliable seed authentication is critical to sustaining rice productivity and economic stability. Portable NIR spectroscopy is a non-destructive, user-friendly, and affordable analytical technique that can rapidly distinguish rice seeds from rice paddy by analyzing their spectral signatures. Miniaturized NIR technology has been increasingly applied in various fields, including agriculture, food, pharmaceuticals, and petrochemicals [14,51,52,6].

This technique works by measuring the absorption of near-infrared light, which varies according to the molecular composition of the sample. When coupled with chemometrics, it provides real-time, actionable information. The ability to determine several quality parameters in real-time without extensive sample preparation or specialized laboratory infrastructure makes portable NIR spectroscopy particularly valuable in preventing seed fraud in Africa and other developing regions. The portability of NIR devices enables onsite testing at farm gates and other agricultural settings, allowing farmers, seed inspectors, and regulatory authorities to quickly and accurately verify rice seed authenticity and quality. This capability is essential in the context of climate change and food security.

Several studies have demonstrated the use of portable NIR spectroscopy for various rice-related applications, including rice quality measurement [10,24,36,39,7], rice freshness identification [30,57,7], quality trait assessment [3], and fatty acid evaluation in rice storage [22,58]. However, there has been little research on using portable NIR spectroscopy for the rapid, non-destructive detection of rice seed integrity, specifically the differentiation of rice seeds from rice paddy to combat fraud along the rice value chain. Additionally, there has been limited discussion on the feasibility of simultaneously determining rice varieties within the two rice categories (seed and paddy).

This study, therefore, explores the application of portable NIR spectroscopy as a tool for differentiating rice seeds from rice paddy in Africa. By providing a rapid, reliable, and accessible method for verifying seed authenticity, NIR spectroscopy can play a crucial role in combating seed fraud, protecting farmers, and supporting the overall integrity of the agricultural supply chain across the continent. The simultaneous determination of rice varieties and seed integrity through portable NIR spectroscopy, combined with chemometrics, offers a user-friendly technique that could significantly enhance rice yields, improve quality control, and contribute to food security.

2. Materials and methods

2.1. Samples acquisition

This study involved four widely cultivated rice varieties in Ghana: Amankwatia, Agra, Jasmine 85, and Legon 1. Certified rice seeds and rice paddy for each variety were sourced from registered suppliers, research stations, and farmers in various regions across Ghana, including Volta, Ashanti, Greater Accra, Central, and Northern regions, during the 2022–2024 seasons. The two main sample categories in this research were rice seeds and rice paddy from the four mentioned varieties. Seventy (70) samples were collected per category for each variety, resulting in 280 samples of rice seeds and 280 samples of rice paddy. Chemical analyses for each variety were derived from prior research studies [13,45], as shown in Table 1.

Table 1
Chemical properties of the samples used (%).

Varieties	Ash	Fat	Protein	Carbohydrate	Fibre	Fe (mg/100 g)	Ca (mg/100 g)
Amankwatia	0.55	1.06	8.00	80.58	0.59	1.60	2.28
Agra	0.56	0.49	8.26	81.19	0.60	0.89	2.45
Jasmine 85	0.61	0.52	5.79	77.87	0.13	0.50	1.10
Legon 1	0.50	0.70	5.80	79.2	0.05	0.53	1.11

2.2. NIR spectra collection

In this study, a portable NIR spectrometer (Innospectra Co., Hsinchu, Taiwan) with a wavelength range of 900–1700 nm was used to acquire spectral data for all samples, including both individual rice seeds and paddy. As per the protocol outlined by Teye and Amuah [45], each sample was scanned individually five times. The spectrometer operated in conjunction with a smartphone (Samsung A95, Android version) via Bluetooth connectivity, enabling real-time data acquisition, visualization, and storage. The spectra was collected in reflectance mode. The collected spectra were then exported for further analysis and classification using chemometric models. This integrated system was commercially available and not independently developed by the authors. The spectrometer's performance metrics included a resolution of 3 nm and spectral responsivity of 10 nm resolution, with an integration time of approximately 1 s per scan. To minimize environmental variations and background noise, a background or dark spectrum was recorded after every 20 sample scans. All measurements were conducted under controlled conditions at 28 °C room temperature and 70 % relative humidity. For sample preparation, individual seeds were scanned rather than seed bunches, ensuring a standardized approach to spectral data collection. All statistical analyses were performed using MATLAB.

2.3. Spectra data processing

In this study, four data pre-processing methods were utilized: multiplicative scatter correction (MSC), standard normal variate (SNV), first derivative (FD), and second derivative (SD). For a detailed explanation of the theory and application of these methods, please refer to our earlier publications and the works of other researchers [32,46,47].

2.4. Principal component analysis (PCA)

Following the pre-processing of the spectral data, Principal Component Analysis (PCA) was performed as the initial analytical step. This allowed us to observe natural clustering patterns within the data and to reduce its dimensionality. PCA is a well-established unsupervised pattern recognition technique used to extract significant information from correlation matrices, facilitating the visualization of data trends in a simplified scatter plot with fewer dimensions. Please refer to our previous publication [46]. In this study, the first three principal components (PCs) captured the most relevant information while minimizing redundancy, ensuring an effective representation of spectral variations. The Kaiser Criterion was applied to retain principal components with eigenvalues greater than 1, ensuring that only the most significant components contributing to data variance were selected for analysis.

2.5. Linear discriminant analysis (LDA)

Linear Discriminant Analysis (LDA) was employed after pre-processing the spectral data to improve the separation between rice varieties and enhance classification accuracy. LDA is a widely used supervised classification method that maximizes the differences between predefined groups, finding the best linear combinations of features that can separate these groups [18,34]. By projecting the data onto a reduced-dimensional space, LDA retains the most important information for distinguishing between classes, allowing for clearer clustering and more reliable classification [21].

To enhance classification accuracy while reducing computational complexity, PCA was first applied as a dimensionality reduction technique. The most significant principal components (PCs) were retained based on the Kaiser Criterion and subsequently used as input variables for the LDA model. LDA, as a supervised classification method, was then applied to maximize class separability and improve the predictive performance of the model.

2.6. Multivariate classification algorithms

Random forest (RF) is an ensemble method based on multiple classification trees for accurate discrimination. It employs variable importance metrics and data resemblance measures for visualization and clustering [2]. K-nearest neighbor (KNN) is a nonparametric algorithm that assesses distances between samples from a calibration set and unknown samples [48]. The parameter K exerts a significant influence on the classification rate of the KNN model. With customized kernel functions, SVM learning methods can imitate complicated non-linear boundaries while exhibiting good performance when it comes to generalization. SVM has been effectively applied in chemometrics to classify near-infrared spectra. It seeks to maximize inter-class geometric margin while minimizing classification error [12]. The optimization of hyperparameters for both the SVM and RF models was performed using grid search tuning to achieve the best classification performance. For the SVM model, the RBF kernel was used with a regularization parameter (C) of 1.0 and gamma set to 0.1. The RF model was optimized with 150 trees (n_estimators), a maximum depth of 10, and a minimum sample per leaf of 3. Model performance was evaluated based on classification accuracy (Eq. 1). NN is a prevalent machine learning approach, leveraging extensive datasets, advanced computing, and sophisticated algorithms.

As a supervised ML technique, NN can approximate any function with a sufficiently large hidden layer. Neural networks excel in capturing complex non-linear relationships among input and output variables, consisting of input, hidden, and output layers [40]. For further details on the models used in this research, consult [29,9].

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{TN + TP + FN + FP} \tag{1}$$

2.7. Data partition

This research employed a comprehensive data validation approach to ensure a dependable model assessment and mitigate overfitting concerns. The spectral information underwent meticulous preprocessing and was formatted appropriately for analysis. A 10-fold cross-validation technique was utilized to verify the model's efficacy. The dataset was divided into 10 equal segments, with each segment serving once as a test set whilst the remaining nine functioned as training data [41]. This procedure was repeated for all ten subsets, guaranteeing that every data point contributed to the training and testing phases. The ultimate model performance metrics were derived from the average across all folds, thoroughly evaluating the model's precision and consistency. By implementing 10-fold cross-validation, the researchers ensured that the outcomes were not skewed by a single data division, thereby enhancing the robustness and generalizability of the results. Fig. 1 illustrates this process.

3. Results and discussion

3.1. Spectral information

Fig. 2 (a1) and (b1) show all the raw reflectance spectra collected for the rice varieties. Fig. 2 (a2) and (b2) present the mean spectral profiles of the four authentic rice varieties. The spectral data obtained in this experiment revealed distinct NIR reflectance spectra features specific to the different categories of rice samples. Fig. 2 present the spectral data for four rice varieties (Agra, Amankwatia, Jasmine 85, and Legon 1), including (a1 and b1) raw rice, (a2 and b2) mean rice.

In Fig. 2a1, notable absorption peaks are visible at approximately 950 nm, 1200 nm, 1400 nm, and 1650 nm. These peaks provide valuable insights into the functional groups present in rice, such as carbohydrates, vitamins, phenols, and minerals. The region between 1000 nm and 1200 nm corresponds to the O-H, C-H combination, and C═O, which are associated with water absorption, hydrocarbons, aromatic, and aliphatic ketone compounds [55,8]. Meanwhile, the bands between 1400 nm and 1650 nm represent the N-H first overtone (related to amide/protein) and the O-H first overtone (related to cellulose, alcohols, and water in rice).

Additionally, Fig. 2a2 highlights characteristic differences among the rice varieties. The spectra indicate that the Amankwatia and Jasmine varieties are closely aligned across the spectral range, suggesting they may share similar properties. Similarly, Agra and Legon 1 exhibit comparable spectral features, indicating potential similarities between these varieties. Spectra data observed in the rice seed were similar to those observed in the rice paddy.

3.2. Principal component analysis and Linear discriminant analysis

Fig. 3 represents the outcomes of principal component analysis and linear discriminant analysis applied to rice seed varieties. Fig. 3a shows how the spectra data of rice is reduced into two dimensions while retaining the most significant variance. The two axes, PC1 (which explains 97.88 % of the variance) and PC2 (which explains 1.925 %), reflect how the data is spread. There is significant overlap among varieties (Agra, Amankwatia, Jasmine 85, and Legon 1), particularly along PC1, which indicates that the most variance in the dataset is shared between the varieties, making it difficult to distinguish them based on this component alone. The

Fig. 1. 10-fold cross-validation.

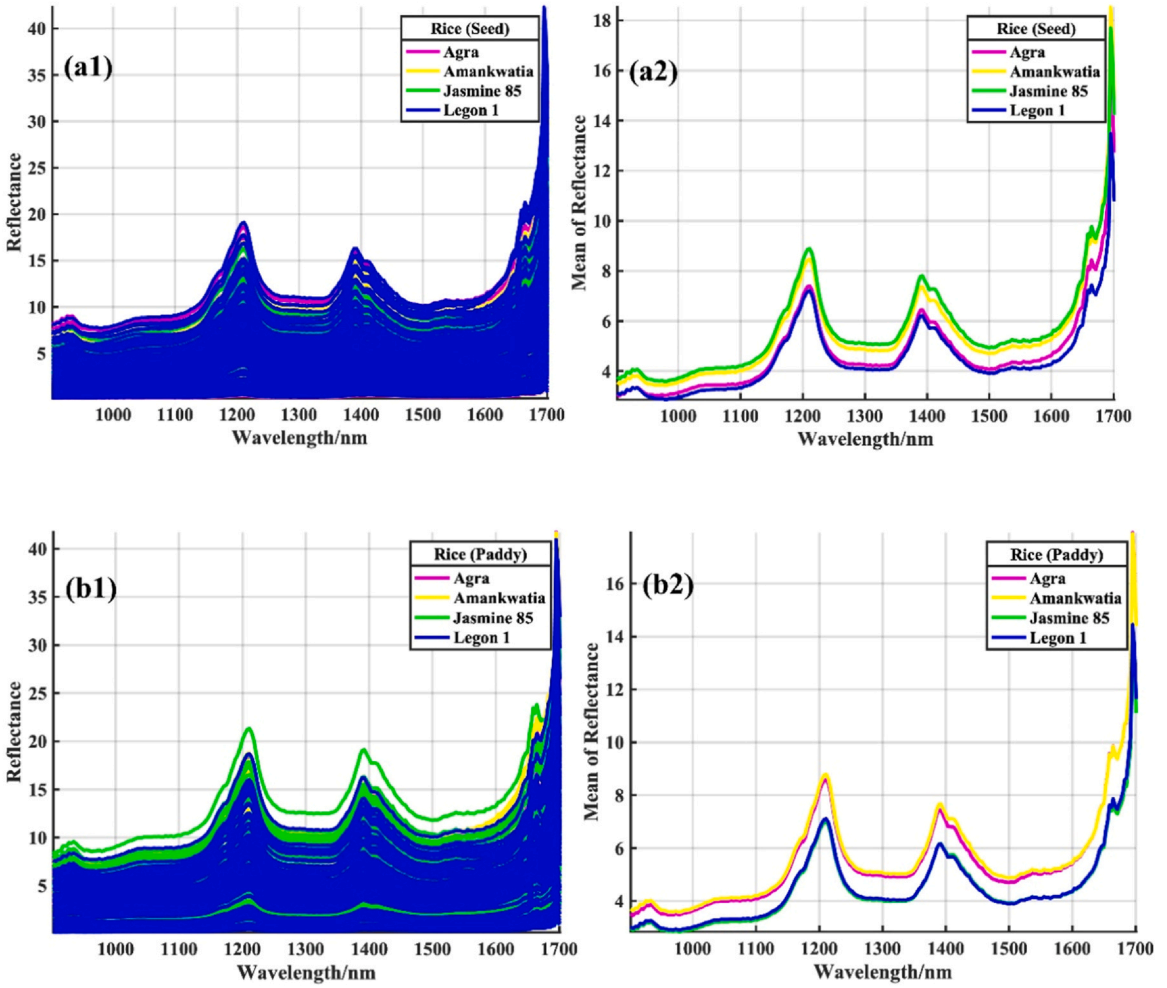

Fig. 2. Spectral information for rice seed samples: (a1) raw spectra and (a2) mean spectra; and for rice paddy samples: (b1) raw spectra and (b2) mean spectra.

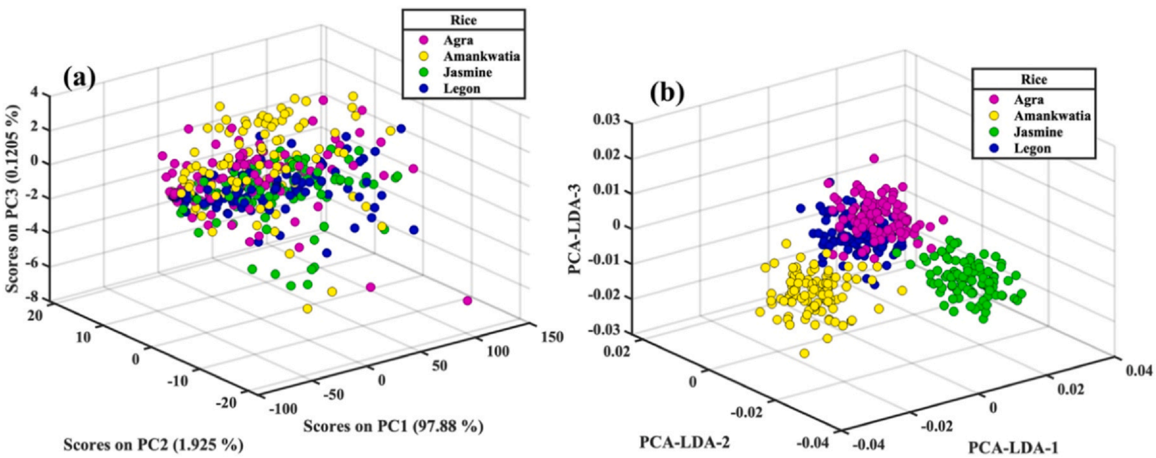

Fig. 3. Score plot (A) PCA and (B) PCA-LDA for rice varieties.

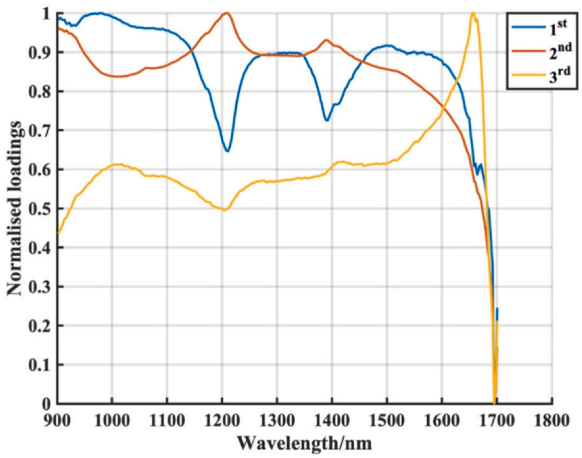

Fig. 4. Loading plots for PCA-LDA for rice varieties.

data points are closely clustered along PC1, with PC2 showing only a minor variation between the varieties. This suggests that while PCA simplifies the dataset by reducing the dimensions, it does not separate the rice varieties.

Fig. 3b, however, shows the separation between the rice varieties based on a supervised classification method. LDA clearly distinguishes the four rice varieties. Although Agra and Legon 1 show some overlap, they are still more distinguishable than the PCA plot. Jasmine 85 and Amankwatia were well-separated, with minimal overlap, indicating that LDA successfully differentiates these varieties. LDA successfully classifies rice seeds based on their spectra data, making it a better tool for identifying and verifying seed varieties. Fig. 4 illustrates the PCA-LDA loading plots for differentiating four rice varieties (Agra, Amankwatia, Jasmine 85, and Legon 1), highlighting key spectral features. For PC 1, major peaks were identified at approximately 937–1033 nm, 1170 nm, 1207 nm, 1389 nm, and 1409 nm. For PC 2, significant peaks were observed at 1170 nm, 1207 nm, 1382 nm, and 1413 nm, while for PC 3, prominent peaks appeared at 929–1055 nm, 1203 nm, and 1413 nm. Previous studies have linked the peak at valley around 1207 nm to the second overtone of C–H stretching vibrations of carbohydrates [54]. Similarly, the absorption band at 1207 nm, attributed to lipids or fat content, originates from the second overtone of C–H stretching mode [53]. These findings emphasize the importance of these spectral regions in distinguishing the rice varieties.

3.3. Principal component analysis and linear discriminant analysis of each rice variety for seed and paddy

In this study, PCA was not used as a standalone classification method but rather as a preprocessing step to extract relevant spectral features before applying LDA. The PCA-reduced dataset was then fed into the LDA model, which optimized class separation and enhanced classification performance. The PCA-LDA score plots for the rice varieties demonstrate a clear cluster separation between the seed and paddy categories compared to only the traditional PCA score plot. As seen in Fig. 5(a-d), these results highlight that combining principal component analysis and linear discriminant analysis effectively distinguished these two categories. The plots show a clear separation between the clusters representing the seeds and paddy for each variety. The two categories are well differentiated in the PCA-LDA space, confirming that there are differences in their spectra data, and hence, it reveals some differences in their chemical properties. Seeds contain higher levels of nutrients and have greater viability, which is essential for plant growth, compared to paddy, which is predominantly composed of carbohydrates and may include impurities and off types. This compositional difference is reflected in the separation between seed and paddy clusters. The ability to distinguish between seeds and paddy is crucial for detecting fraudulent practices, where lower-quality paddy may be mixed with higher-quality seeds to cut costs. This method can ensure that only pure, high-quality seeds are distributed to farmers, protecting them from receiving substandard planting materials. The separation also underscores the economic value of maintaining seed purity. Given that seeds are more expensive due to their superior quality, precise identification is vital to prevent financial losses caused by the introduction of paddy into seed lots.

3.4. PCA and PCA-LDA for general cluster trends

A global cluster trend was attempted for several varieties in this study to provide evidence that certified seeds could be identified easily from paddy, irrespective of the variety. As observed in Fig. 6, the classification of rice seeds and paddy using a combination of Principal Component Analysis and Linear Discriminant Analysis revealed some helpful information when several varieties were used. The plot shows that this combined method effectively separates rice seeds from paddy, although some overlap remains. This overlap indicates that, in some cases, the spectral data from seeds and paddy share similarities though only a few samples showed that. This phenomenon could be attributed to a few paddy grains that show variability. Despite this, the PCA-LDA technique provides valuable insight and is helpful for practical classification, especially in scenarios like seed quality verification or detecting possible mixing of seeds and paddy. Fig. 7 presents the PCA-LDA loading plot for rice seed and paddy. The loadings emphasize the spectral regions that differentiate between seeds and paddy, driven by their distinct chemical properties. Major peaks were observed at 926–1028 nm,

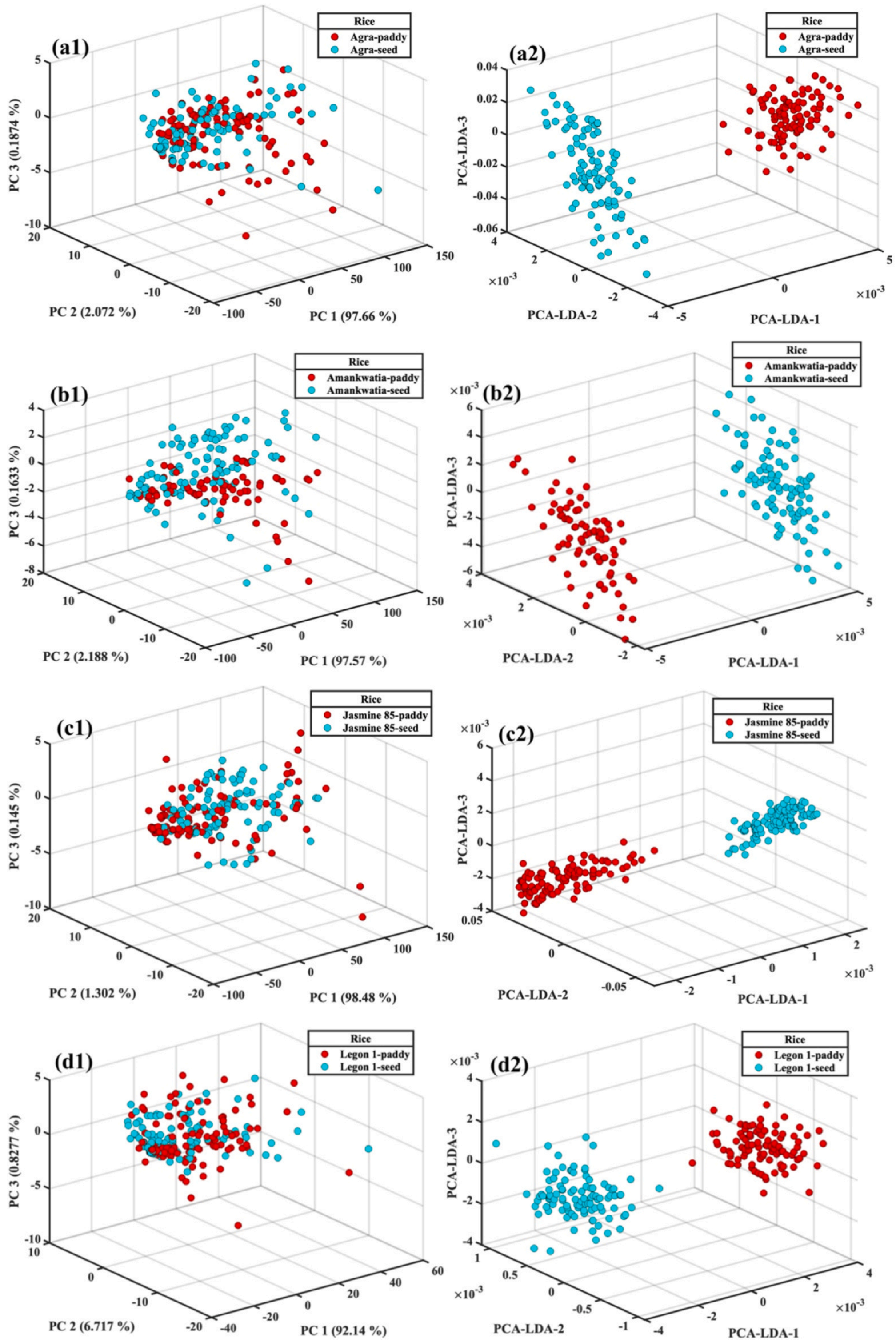

Fig. 5. PCA and PCA-LDA score plot of rice samples used (seeds and paddy): (a) Agra, (b) Amankwatia, (c) Jasmine 85, and (d) Legon 1.

1170 nm, 1207 nm, 1389 nm, 1409 nm, and 1436–1646 nm for PC 1. For PC 2, significant peaks were identified at 933–1139 nm, 1207 nm, 1392 nm, and 1409 nm, while for PC 3, major peaks were observed at 933–1060 nm, 1203 nm, and 1406 nm. The PCA loadings revealed that the key wavelengths contributing to the spectral variance among rice varieties were primarily associated with carbohydrate, and moisture content, which are major chemical components influencing NIR absorption [31]. The region

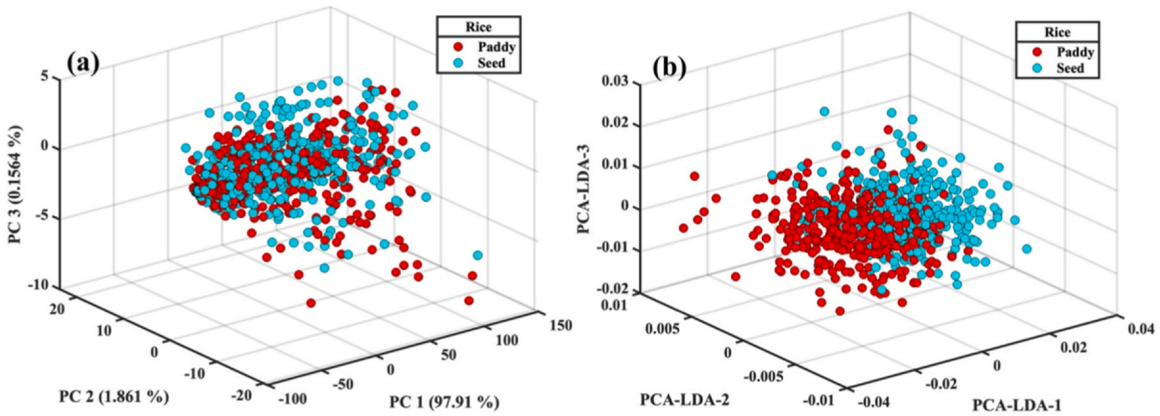

Fig. 6. PCA and PCA-LDA score plot of rice seed and rice paddy from four varieties.

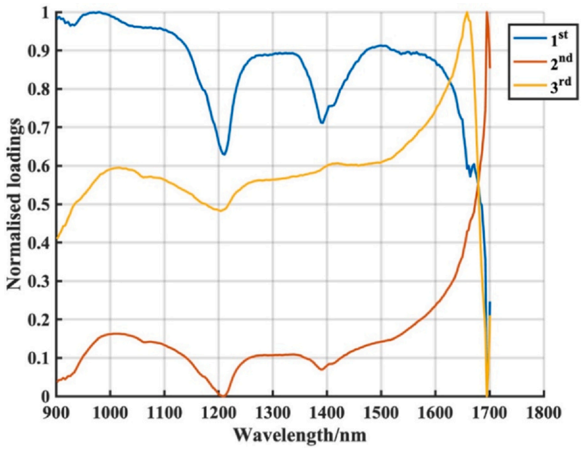

Fig. 7. PCA-LDA loading plot of rice seed and rice paddy from four varieties.

926–1028 nm is associated with the N-H second overtone, 1170 nm corresponds to the C-H second overtone, and 1436–1646 nm is linked to the O-H and N-H first overtones [15]. Other researchers have also reported some of these peaks in rice samples [45].

3.5. Classification of varieties (paddy) and (seed)

The research further explored the classification of varieties under seed and paddy categories differently using three pre-processing techniques compared with raw, as seen in Table 2. It was observed that each pre-processing technique positively impacted the performance of the classification model, similar to those observed by other researchers [46]. For the seed classification, the optimum performance was 100 % and 99 % for both the training and test sets, respectively, for MSC-RF and FD-RF as seen in Fig. 8. The table showed that RF performed better than the others in all the pre-processing techniques employed, with general performance above 92 % (Fig. 8). More specifically, the best model was MC-RF, which gave an optimum performance of 100 % and 97.63 % at the training and test sets, respectively, for the paddy. It could be explained that the properties of Random Forest can handle several input variables without variable deletion and generate an internal unbiased estimate of the generalization error as the forest building progresses without overfitting and it is known to be unexcelled in accuracy among current algorithms.

3.6. Classification of rice seeds from rice paddy

Accurate and rapid identification of rice seeds from rice paddy or fake rice seeds would present a novel detection that could promote rice productivity and enhance food security. Table 3 presents the outcome of several pre-processing techniques and a multivariate algorithm for model development. The results show that, for a single variety, the model could accurately differentiate authentic rice seed from paddy or fake rice seed. The RF and SVM models gave a 100 % classification rate in training and test sets, signifying those authentic seeds could be accurately identified from fake seeds to support farmers in sustainable rice production. The model's performance was not different from those reported by other researchers [1,45]. While the classification models achieved 100 % accuracy for most rice varieties, the Jasmine variety exhibited lower performance, indicating potential spectral similarities

Table 2
Performance of the classification models for authenticating the rice seed and rice paddy.

Types	Models	Pre-processing											
		RAW		MC		MSC		SNV		FD		SD	
		Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
Paddy	K-NN	98.66	97.91	98.05	96.59	98.48	96.86	97.81	95.80	88.98	79.02	93.76	88.72
	NN	99.36	96.07	99.91	96.08	100.00	96.33	99.74	95.80	90.38	84.79	97.32	92.40
	RF	**100.00**	**97.38**	**100.00**	**97.13**	**100.00**	**96.85**	**100.00**	**96.85**	**100.00**	**80.31**	**100.00**	**92.40**
	LDA	97.17	96.59	97.26	97.38	97.32	96.86	97.84	97.65	87.66	87.16	93.29	93.44
	QDA	97.87	97.12	97.08	96.60	97.84	97.12	98.02	96.86	88.19	86.63	95.01	93.70
	SVM	97.14	96.33	96.65	95.28	97.87	96.34	97.61	97.38	87.05	86.10	93.85	92.66
Seed	K-NN	99.56	98.25	98.56	97.75	98.92	97.50	99.33	97.00	86.28	74.25	97.14	95.50
	NN	100.00	97.25	100.00	97.50	100.00	99.00	100.00	100.00	96.28	93.00	99.86	93.00
	RF	**100.00**	**97.00**	**100.00**	**97.50**	**100.00**	**98.50**	**100.00**	**99.25**	**100.00**	**72.00**	**100.00**	**94.50**
	LDA	98.61	98.25	98.11	98.00	99.50	99.50	99.75	99.75	95.25	94.75	95.97	96.00
	QDA	98.97	98.50	98.89	98.25	99.53	99.50	99.86	99.50	95.39	95.00	96.64	96.25
	SVM	99.06	98.00	99.11	98.00	98.94	99.00	99.69	99.50	89.47	89.00	96.83	96.25

Note: bold figures were the best

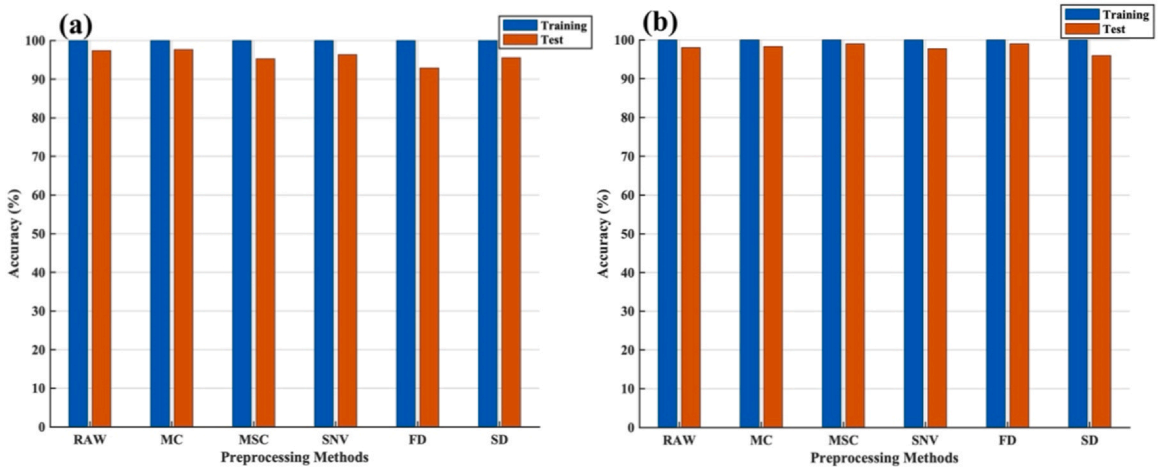

Fig. 8. Pre-processing methods and performance of the Random Forest (RF) model on paddy (a) and Seed (b).

Table 3
Performance of the classification models for authenticating the rice seed and rice paddy.

Model	**Agra (seed and paddy/fake seed) - Pre-processing**							
	RAW		MC		FD		SD	
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
LDA	51.41	41.29	52.57	43.29	100.00	100.00	100.00	100.00
QDA	70.32	70.17	61.97	58.69	100.00	100.00	100.00	100.00
RF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SVM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Model	**Amankwatia (seed and paddy/fake seed) Pre-processing**							
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
LDA	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
QDA	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
RF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SVM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Model	**Legon1 (seed and paddy/fake seed) Pre-processing**							
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
LDA	50.33	45.50	50.83	37.00	50.83	43.50	52.00	40.00
QDA	51.67	46.00	57.56	50.00	57.78	55.00	54.17	44.00
RF	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
SVM	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Model	**Jasmine (seed and paddy/fake seed) Pre-processing**							
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
LDA	50.72	61.00	51.44	36.00	50.00	50.00	50.00	50.00
QDA	51.22	50.50	61.61	57.00	50.00	50.00	50.00	50.00
RF	50.00	50.00	100.00	100.00	50.00	50.00	100.00	100.00
SVM	50.00	50.00	100.00	100.00	50.00	50.00	50.00	50.00

with other varieties. This may be attributed to overlapping chemical properties, particularly in starch, which affect NIR spectral absorption.

Furthermore, for a generalized model, this research further attempted to classify rice seed and paddy irrespective of the type of varieties (as in this case, four varieties of seed and paddy in this study) for seed authenticity, as seen in Table 4, a system to check rice seed fraud. For classifying the two categories (seed and paddy), FD-LDA gave 83.33 % in the training set and 83.35 % in the test set (Fig. 9). Though the accuracy was below 90 %, the model provides a feasibility or potential for rapid identification of rice seed fraud at Farmgate. These results show that the models could be used for the rapid screening of rice seeds and could assist in market surveillance to ensure seed integrity along the agricultural value chain. While this study provides a proof-of-concept, future research should expand the dataset to include a wider range of rice varieties, particularly those with minimal genetic differences

Table 4
Performance of classification model for rice seed and rice paddy from four varieties.

Models	**Rice seed and rice paddy / fake seed - Pre-processing**											
	RAW		MC		MSC		SNV		FD		SD	
	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
LDA	**81.89**	**81.83**	**81.79**	**81.57**	**82.70**	**82.72**	**82.89**	**82.84**	**83.33**	**83.35**	**78.38**	**78.36**
QDA	80.42	79.65	80.92	80.79	81.25	82.07	80.58	79.00	83.17	82.84	55.56	55.44
RF	51.22	51.22	51.22	51.22	51.22	51.22	51.22	51.22	100.00	80.28	51.22	51.22
SVM	51.22	51.22	51.22	51.22	51.22	51.22	51.22	51.22	83.35	83.35	51.22	51.22

Note: bold figures were the best

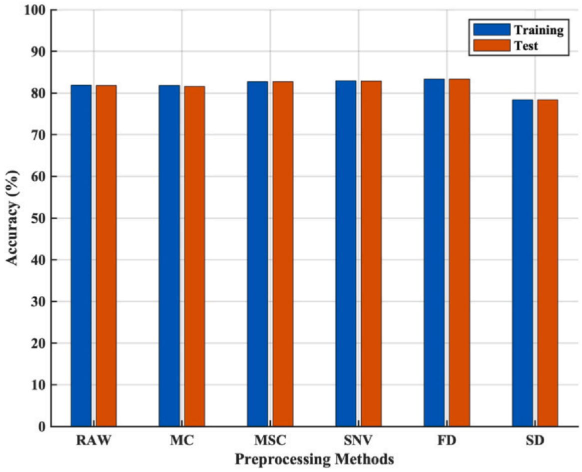

Fig. 9. Performance of the Linear Discriminant Analysis (LDA) model for rice seed and rice paddy across different preprocessing techniques.

4. Conclusion

The research outcome provides the solution for checking rice seed fraud in real-time onsite to promote food integrity. The results show that among the four classification algorithms used, RF gave an accuracy of 99.0 % and 95.28 % for rice seed and paddy varietal identification, respectively, in the test sets. For the authentication of rice seed and paddy to check seed fraud, SVM gave 98.15 % accuracy in the test set. This shows that a portable NIR device could be employed for simultaneous rice varietal and seed integrity determination along the value chain. This technology would be helpful for seed inspectors, farmers, and regulatory officers in the rice industry.

CRediT authorship contribution statement

Ernest Teye: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Charles Lloyd Yeboah Amuah:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Vida Gyimah Boadu:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kwadwo Anokye Dompreh:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Maxwell Darko Asante:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Francis Padi Lamptey:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stephen Narh:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Daniel Dzorkpe Gamenyah:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **George Oduro Nkansah:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Selorm Akaba:** Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the support provided by the Directorate of Research, Innovation and Consultancy (RSG/INT/CANS/2023/101) of the University of Cape Coast. We also express our sincere gratitude to the African Agribusiness Consortium of the Jospong Group of Companies for generously providing the rice seeds used in this study. The assistance provided by Mr. Joel Welbeck and Mrs Winifred Dogbe is highly acknowledged.

References

[1] E.K. Anyidoho, E. Teye, R. Agbemafle, Nondestructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm, Anal. Methods 12 (33) (2020) 4150–4158.

[2] E.K. Anyidoho, E. Teye, R. Agbemafle, Differentiation of organic cocoa beans and conventional ones by using handheld NIR spectroscopy and multivariate classification techniques, Int. J. Food Sci. 2021 (1) (2021) 1844675, https://doi.org/10.1155/2021/1844675.

[3] A. Aznan, C. Gonzalez Viejo, A. Pang, S. Fuentes, Rapid assessment of rice quality traits using low-cost digital technologies, Foods 11 (9) (2022) 1181.

[4] V. Balasubramanian, M. Sie, R. Hijmans, K. Otsuka, Increasing rice production in sub-Saharan Africa: challenges and opportunities, Adv. Agron. 94 (2007) 55–133.

[5] N. BauerCheating the senses. Paper presented at the Proceedings of the Oxford Symposium on Food and Cookery.2020.

[6] K.B. Beć, J. Grabska, H.W. Siesler, C.W. Huck, Handheld near-infrared spectrometers: where are we heading? NIR N. 31 (3-4) (2020) 28–35.

[7] Y.-K. Chuang, Y.-P. Hu, I.-C. Yang, S.R. Delwiche, Y.M. Lo, C.-Y. Tsai, S. Chen, Integration of independent component analysis with near infrared spectroscopy for evaluation of rice freshness, J. Cereal Sci. 60 (1) (2014) 238–242.

[8] J. Coates, Interpretation of infrared spectra, a practical approach, Encycl. Anal. Chem. 12 (2000) 10815–10837.

[9] V. Cortés, J.M. Barat, P. Talens, J. Blasco, M.J. Lerma-García, A comparison between NIR and ATR-FTIR spectroscopy for varietal differentiation of Spanish intact almonds, Food Control 94 (2018) 241–248.

[10] M.C. Custodio, R.P. Cuevas, J. Ynion, A.G. Laborte, M.L. Velasco, M. Demont, Rice quality: How is it defined by consumers, industry, food scientists, and geneticists? Trends Food Sci. Technol. 92 (2019) 122–137.

[11] V.D. Daygon, S. Prakash, M. Calingacion, A. Riedel, B. Ovenden, P. Snell, M. Fitzgerald, Understanding the Jasmine phenotype of rice through metabolite profiling and sensory evaluation, Metabolomics 12 (2016) 1–15.

[12] O. Devos, C. Ruckebusch, A. Durand, L. Duponchel, J.-P. Huvenne, Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation, Chemom. Intell. Lab. Syst. 96 (1) (2009) 27–33.

[13] C. Diako, E. Sakyi-Dawson, B. Bediako-Amoa, F. Saalia, J. Manful, Cooking characteristics and variations in nutrient content of some new scented rice varieties in Ghana, Ann. Food Sci. Technol. 12 (1) (2011) 39–44.

[14] C.A.T. Dos Santos, M. Lopo, R.N. Páscoa, J.A. Lopes, A review on the applications of portable near-infrared spectrometers in the agro-food industry, Appl. Spectrosc. 67 (11) (2013) 1215–1233.

[15] A.B. Eldin, I. Akyar, Near infra red spectroscopy, Wide Spectra of Quality Control, InTech, *Rijeka, Croatia*, 2011, pp. 237–248.

[16] A.C. Faller, P. Kesanakurti, T. Arunachalam, Fraud in grains and cereals, Food Fraud, Elsevier, 2021, pp. 281–308.

[17] R. Fiamohe, M. Demont, K. Saito, H. Roy-Macauley, E. Tollens, How can West African rice compete in urban markets? A demand perspective for policymakers, EuroChoices 17 (2) (2018) 51–57.

[18] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (2) (1936) 179–188.

[19] A. Haque, M. Akon, M. Islam, K. Khalequzzaman, M. Ali, 2007, Study of seed health, germination and seedling vigor of farmers produced rice seeds..

[20] R. Harisha, D. Bhadru, S. Vanisri, V. Gourishanakar, L. Satish, SSR and morphological traits based fingerprints and DNA barcodes for varietal identification in rice, Biotechnol. Biotechnol. Equip. 35 (1) (2021) 1461–1473.

[21] T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, 2009 (2).

[22] H. Jiang, T. Liu, Q. Chen, Dynamic monitoring of fatty acid value in rice storage based on a portable near-infrared spectroscopy system, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 240 (2020) 118620.

[23] H. Jung, S.-J. Lee, J.H. Lim, B.K. Kim, K.J. Park, Chemical and sensory profiles of makgeolli, Korean commercial rice wine, from descriptive, chemical, and volatile compound analyses, Food Chem. 152 (2014) 624–632.

[24] S. Kawamura, M. Natsuga, K. Takekura, K. Itoh, Development of an automatic rice-quality inspection system, Comput. Electron. Agric. 40 (1-3) (2003) 115–126.

[25] T. Kawasaki, K. Mizuno, T. Baba, H. Shimada, Molecular analysis of the gene encoding a rice starch branching enzyme, Mol. Gen. Genet. 237 (1993) 10–16.

[26] T. Kleawphaipan, S. Somprasong, T. Srahongthong, B. PattanasiriThai Hom Mali rice purity test by using digital image analysis. Paper presented at the Journal of Physics: Conference Series.2019.

[27] S. Komatsu, H. Konishi, S. Shen, G. Yang, Rice proteomics: a step toward functional analysis of the rice genome, Mol. Cell. Proteom. 2 (1) (2003) 2–10.

[28] K. Kuhlmann, T. Francis, I. Thomas, 2021. Seed laws and regulations affecting the development of the private vegetable seed sector in Sub-Saharan Africa. In: Shanhua..

[29] F.P. Lamptey, E. Teye, E.E. Abano, C.L. Amuah, Application of handheld NIR spectrometer for simultaneous identification and quantification of quality parameters in intact mango fruits, Smart Agric. Technol. 6 (2023).

[30] H. Lin, H. Jiang, J. Lin, Q. Chen, S. Ali, S.W. Teng, M. Zuo, Rice freshness identification based on visible near-infrared spectroscopy and colorimetric sensor array, Food Anal. Methods 14 (2021) 1305–1314.

[31] L. Lin, Y. He, Z. Xiao, K. Zhao, T. Dong, P. Nie, Rapid-detection sensor for rice grain moisture based on NIR spectroscopy, Appl. Sci. 9 (8) (2019) 1654.

[32] J. Luypaert, S. Heuerding, Y. Vander Heyden, D. Massart, The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams, J. Pharm. Biomed. Anal. 36 (3) (2004) 495–503.

[33] M. Manohar, K. Chatrapathy, M. Sowmya, 2017. Smart detection of rice purity and its grading. Paper presented at the 2017 3rd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)..

[34] G.J. McLachlan, Discriminant Analysis and Statistical Pattern Recognition, John Wiley & Sons, 2005.

[35] T.T. Nguyen, V.-N. Hoang, T.-L. Le, T.-H. Tran, H. Vu, 2018. A vision based method for automatic evaluation of germination rate of rice seeds. Paper presented at the 2018 1st International Conference on Multimedia Analysis and Pattern Recognition (MAPR)..

[36] Y. Ogawa, Quality evaluation of rice, Computer Vision Technology for Food Quality Evaluation, Elsevier, 2016, pp. 413–437.

[37] C.M. Paule, J. Powers, Sensory and chemical examination of aromatic and nonaromatic rices, J. Food Sci. 54 (2) (1989) 343–346.

[38] R. Prasad, Y.S. Shivay, D. Kumar, Current status, challenges, and opportunities in rice production, Rice Prod. Worldw. (2017) 1–32.

[39] C. Prom-U-Thai, B. Rerkasem, Rice quality improvement. A review, Agron. Sustain. Dev. 40 (4) (2020) 28.

[40] X. Qi, G. Chen, Y. Li, X. Cheng, C. Li, Applying neural-network-based machine learning to additive manufacturing: current applications, challenges, and future perspectives, Engineering 5 (4) (2019) 721–729.

[41] I. Ramírez-Morales, D. Rivero, E. Fernández-Blanco, A. Pazos, Optimization of NIR calibration models for multiple processes in the sugar industry, Chemom. Intell. Lab. Syst. 159 (2016) 45–57, https://doi.org/10.1016/j.chemolab.2016.10.003.

[42] W. Ramakishana, M. Lagu, V. Gupta, P. Ranjekar, DNA fingerprinting in rice using oligonucleotide probes specific for simple repetitive DNA sequences, Theor. Appl. Genet. 88 (1994) 402–406.

[43] V. Satturu, D. Rani, S. Gattu, J. Md, S. Mulinti, R.K. Nagireddy, R. Yanda, DNA fingerprinting for identification of rice varieties and seed genetic purity assessment, Agric. Res. 7 (4) (2018) 379–390.

[44] M. Śliwińska-Bartel, D.T. Burns, C. Elliott, Rice fraud a global problem: a review of analytical tools to detect species, country of origin and adulterations, Trends Food Sci. Technol. 116 (2021) 36–46.

[45] E. Teye, C.L. Amuah, Rice varietal integrity and adulteration fraud detection by chemometrical analysis of pocket-sized NIR spectra data, Appl. Food Res. 2 (2) (2022) 100218.

[46] E. Teye, C.L. Amuah, T. McGrath, C. Elliott, Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 217 (2019) 147–154.

[47] E. Teye, X. Huang, H. Dai, Q. Chen, Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 114 (2013) 183–189.

[48] P. Thanh Noi, M. Kappas, Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery, Sensors 18 (1) (2017) 18.
[49] N. Van Nguyen, A. Ferrero, Meeting the challenges of global rice production, Paddy Water Environ. 4 (2006) 1–9.
[50] F. Vidotto, A. Ferrero, Germination behaviour of red rice (Oryza sativa L.) seeds in field and laboratory conditions, Agronomie 20 (4) (2000) 375–382.
[51] R. Vitorino, A.S. Barros, S. Guedes, D.C. Caixeta, R. Sabino-Silva, Diagnostic and monitoring applications using near infrared (NIR) spectroscopy in cancer and other diseases, Photodiagn. Photodyn. Ther. 42 (2023) 103633.
[52] M. Wang, Y. Xu, Y. Yang, B. Mu, M.A. Nikitina, X. Xiao, Vis/NIR optical biosensors applications for fruit monitoring, Biosens. Bioelectron. X 11 (2022) 100197.
[53] L. Wimonsiri, P. Ritthiruangdej, S. Kasemsumran, N. Therdthai, W. Chanput, Y. Ozaki, Rapid analysis of chemical composition in intact and milled rice cookies using near infrared spectroscopy, J. Infrared Spectrosc. 25 (5) (2017) 330–337.
[54] N. Wu, H. Jiang, Y. Bao, C. Zhang, J. Zhang, W. Song, F. Liu, Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions, Sens. Actuators B Chem. 308 (2020) 127696.
[55] L. Weyer, S. Lo, Spectra-structure correlations in the near-infrared, Handb. Vib. Spectrosc. 3 (2002) 1817–1837.
[56] X. Ye-Yun, Z. Zhan, X. Yi-Ping, Y. Long-Ping, Identification and purity test of super hybrid rice with SSR molecular markers, Chin. J. Rice Sci. 19 (2) (2005) 95.
[57] C. Zhai, W. Wang, M. Gao, X. Feng, S. Zhang, C. Qian, Rapid classification of rice according to storage duration via near-infrared spectroscopy and machine learning, Talanta Open 10 (2024) 100343.
[58] Z. Zhou, C. Blanchard, S. Helliwell, K. Robards, Fatty acid composition of three rice varieties following storage, J. Cereal Sci. 37 (3) (2003) 327–335.