

# Multispectral and X-ray images for characterization of *Jatropha curcas* L. seed quality

**Vitor J Bianchini**

Universidade de Sao Paulo Escola Superior de Agricultura Luiz de Queiroz

**Gabriel M Mascarin**

Empresa Brasileira de Pesquisa Agropecuaria

**Lúcia CAS Silva**

Universidade de Sao Paulo Centro de Energia Nuclear na Agricultura

**Valter Arthur**

Universidade de Sao Paulo Centro de Energia Nuclear na Agricultura

**Jean M Carstensen**

Danmarks Tekniske Universitet

**Birte Boelt**

Aarhus Universitet

**Clíssia Barboza da Silva** (✉ [clissia\\_usp@hotmail.com](mailto:clissia_usp@hotmail.com))

Universidade de Sao Paulo Centro de Energia Nuclear na Agricultura <https://orcid.org/0000-0002-7279-3260>

---

## Methodology

**Keywords:** *Jatropha curcas*, non-invasive methods, radiographic images, artificial intelligence

**Posted Date:** October 26th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-28449/v3>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Plant Methods on January 26th, 2021. See the published version at <https://doi.org/10.1186/s13007-021-00709-6>.

# Abstract

**Background:** The use of non-destructive methods with less human interference is of great interest in agricultural industry and crop breeding. Modern imaging technologies enable the automatic visualization of multi-parameter for characterization of biological samples, reducing subjectivity and optimizing the analysis process. Furthermore, the combination of two or more imaging techniques has contributed to discovering new physicochemical tools and interpreting datasets in real time.

**Results:** We present a new method for automatic characterization of seed quality based on the combination of multispectral and X-ray imaging technologies. We proposed an approach using X-ray images to investigate internal tissues because seed surface profile can be negatively affected, but without reaching important internal regions of seeds. An oilseed plant (*Jatropha curcas*) was used as a model species, which also serve as a multi-purposed crop of economic importance worldwide. Our studies included the application of a normalized canonical discriminant analyses (nCDA) algorithm as a supervised transformation building method to obtain spatial and spectral patterns on different seedlots. We developed classification models using reflectance data and X-ray classes based on linear discriminant analysis (LDA). The classification models, individually or combined, showed high accuracy (>0.96) using reflectance at 940 nm and X-ray data to predict quality traits such as normal seedlings, abnormal seedlings and dead seeds.

**Conclusions:** Multispectral and X-ray imaging have a strong relationship with seed physiological performance. Reflectance at 940 nm and X-ray data can efficiently predict seed quality attributes. These techniques can be alternative methods for rapid, efficient, sustainable and non-destructive characterization of seed quality in the future, overcoming the intrinsic subjectivity of the conventional seed quality analysis.

## Background

A crucial step for crop success is the use of high-quality seeds for obtaining vigorous and uniform seedlings. High-quality seeds are more resistant to biotic and abiotic stresses [1], and they originate seedlings that provide rapid soil coverage, using more efficiently the available radiation and nutrients, reducing the potential side-effects caused by weed-crop competition [2].

Currently, quality monitoring operations are estimated mainly by visual inspection of seeds one by one using germination and vigor tests, which are destructive, laborious and requiring trained seed analysts. New and refined technologies based on computerized procedures create the opportunity for automating laboratory evaluations, providing decision-making support regarding the destination of seedlots. In addition, the development of fast, non-invasive and less subjective tools is relevant for seed industry.

Multispectral imaging can provide spatial and spectral information related to different quality traits, such as surface structure, texture and chemical composition [3]. Briefly, this technique is based on sequential exposure of an object to light at different wavelengths integrated with computer systems to recognize

physicochemical parameters from biological samples. In the context of seed quality, spectral imaging is mainly based on texture [4], physical and chemical attributes associated with insect damages [5], fungal infections [6], among others, but with limitation to examine internal tissues. Meanwhile, X-ray imaging has proved great potential to collect data from internal structures (e.g. damages in embryo and endosperm) [7, 8], because X-rays are short electromagnetic waves (ranging from 0.01 to 10 nm), with high penetrating power [9]. The correlation of these technologies with data obtained by traditional analytical methods can generate new quality markers [10]. A particular combination of spectral data with X-ray imaging can also improve the performance of classifiers [11], providing complementary information related to morphological and biochemical parameters.

In this study, we tested multispectral and X-ray imaging compared to conventional analytical methods for seed quality characterization. *Jatropha curcas* was used as model species, which has been a focus in the study of plants that can be used to produce biofuel, food, feed and biogas from seed cakes [12 – 16]. To the best of our knowledge, there has been no attempts in using multispectral imaging combined with radiographic images prior to this study.

## Results

### Seed vigor classification based on traditional tests

Overall, seeds from Lot 2 showed the best performance in the germination tests (Table 1). Lot 1 had the lowest germination rate, and Lot 3 showed an intermediate performance (germination on paper substrate). In the vigor tests, Lot 2 was classified as higher vigor, with the germination rate index – GRI having greater sensibility in separating seedlots based on vigor: low vigor = Lot 1, high vigor = Lot 2, and medium vigor = Lot 3 (Table 2).

**Table 1 Germination tests to rank *Jatropha curcas* seedlots based on germination capacity (normal seedlings) at 7 and 10 days, using two different substrates (paper and sand).**

Seedlot	Germination [%]		Germination [%]	
	5 days		10 days	
	Paper	Sand	Paper	Sand
1	13 ± 3.00 c*	59 ± 5.25 b	18 ± 2.90 c	63 ± 5.97 b
2	87 ± 3.00 a	94 ± 1.15 a	98 ± 1.33 a	97 ± 1.00 a
3	58 ± 5.54 b	70 ± 2.58 b	60 ± 5.58 b	70 ± 2.58 b
ANOVA	F <sub>2,27</sub> = 85.70	F <sub>2,9</sub> = 26.94	F <sub>2,27</sub> = 116.22	F <sub>2,9</sub> = 22.31
	P = 0.0000	P = 0.002	P = 0.0000	P = 0.0003

\*Means (+/- standard error) within each column followed by the same letter are not significantly different according to Tukey test ( $P < 0.05$ ).

Table 2 Vigor tests to rank *Jatropha curcas* seedlots based on vigor: germination rate index, electrical conductivity and seedling emergence at 10 days.

Seedlot	Germination rate index	Electrical conductivity [ $\mu\text{S}\cdot\text{cm}^{-1}\cdot\text{g}^{-1}$ ]	Seedling emergence [%]
1	0.6 ± 0.49 c	188.8 ± 7.88 b	66 ± 4.76 b*
2	4.0 ± 0.06 a	79.1 ± 3.69 a	91 ± 3.00 a
3	2.5 ± 0.17 b	91.8 ± 5.61 a	59 ± 9.29 b
	$F_{2,27} = 81.31$	$F_{2,9} = 100.71$	$F_{2,9} = 7.19$
ANOVA	$P = 0.0000$	$P = 0.0000$	$P = 0.0136$

\*Means (+/- standard error) within each column followed by the same letter are not significantly different according to Tukey test ( $P < 0.05$ ).

Crude fat content greatly varied among seedlots ( $F = 15.37$ ,  $df = 2, 6$ ,  $P = 0.004$ ), with Lot 2 exhibiting 51% and 35% more fat than Lot 1 and Lot 3, respectively (Fig. 1). Conversely, the crude protein content, known to have a marked effect on the rapid imbibition of water by seeds, did not significantly differ among the seedlots ( $F = 2.03$ ,  $df = 2, 6$ ,  $P = 0.212$ ) (Fig. 1).

### Multispectral imaging integrated with X-ray imaging

Seed orientation was not relevant to discriminate seedlots based on multispectral reflectance (Fig. 2). The wavelengths from ultraviolet (365 nm) and visible (405-690 nm) regions showed low reflectance intensity (<20%) and it was difficult to distinguish the lots. However, data obtained at longer wavelengths, particularly in the near infrared (NIR) region (from 780 to 970 nm) clearly enabled discrimination among seedlots, and seeds with high vigor showed the lowest reflectance intensity (Lot 2).

A principal component analysis (PCA) was applied to the multispectral data to reduce variables and it revealed the first two principal components accounted for most of the data variability (98.66% of total data variance) among seedlots. The contribution histogram indicated the most informative wavelengths based on 95% confidence threshold (cut-off) represented by the dashed red line, which revealed only five bandwidths from 780 to 970 nm that mostly contributed to PC1 and PC2 (Fig. 3A). Cluster analysis based on PC1 and PC2 distinctly separated the three seedlots (Fig. 3B).

In order to validate the PCA model, multispectral data corresponding to the five most important wavelengths (780, 850, 880, 940 and 970 nm), as previously assigned by PCA were used in a canonical discriminant analysis (CDA) (Fig. 4). Lots 1 and 3 had lower vigor, and those wavelengths positively distinguished Lots 1 and 3 from Lot 2 (high vigor), which suggests that these wavelengths are good predictors for discriminating *jatropha* seedlots with lower vigor.

*Jatropha* seed has a thick and dark tegument. Fig. 5 shows RGB images obtained from ventral and dorsal surface of the seeds based on germination capacity and corresponding reflectance images captured at

940 nm and X-ray images. Since the CDA model showed collinearity between 940 and 970 nm (Fig. 4), we selected the reflectance images captured at 940 nm (Fig. 5) because this wavelength is strongly associated with absorbance peak of fatty tissues, which is relevant for oilseed studies.

In the germination test, normal seedlings were produced from seeds with lower pixel values in the reflectance images and higher pixel values in the X-ray images. There was a different pattern from those seeds that generated abnormal seedlings or did not germinate (dead seeds) in which dead seeds exhibited the highest and lowest pixel values in the reflectance and X-ray images, respectively.

Radiographic images obtained from all seedlots were separated into three different classes (Fig. 6A) based on seed tissue integrity and seed performance in the germination test. Soft tissues are associated with deteriorated tissues and they absorb the X-ray beam less as it passed through the tissue, therefore, these areas appear dark in the radiographic images. Meanwhile, regions with high gray intensity indicate greater penetration of X-rays directly associated with higher tissue density (healthy tissues). In class 1, seeds were completely filled or with slight empty spaces ( $\leq 1.23$  mm) between the endosperm and the seed coat, and these seeds mainly generated normal seedlings. Seeds in class 2 showed large empty spaces ( $>1.24$  mm) between the endosperm and the seed coat or deteriorated tissues without reaching the embryonic axis, and they produced mostly abnormal seedlings. Finally, class 3 included seeds with deteriorated tissues reaching the embryonic axis, malformed and empty seeds, which were directly associated with dead seeds.

The CDA method (Fig. 6B) showed that Lot 1 was positively correlated with class 2 and dead seeds. The abnormal seedling variable did not appear in the model because it was collinear with dead seeds. Seeds of Lot 2 were positively correlated with X-ray images in class 1 and normal seedlings, and it was completely opposite to Lots 1 and 3. Reflectance data at 940 nm distinguished Lot 3 from Lots 1 and 2, with higher amount of seeds in class 3 and dead seeds ( $F = 10.22$ ,  $df = 2, 12$ ,  $P = 0.0014$ ).

## Validation

We created models to validate the proposed approach using a Linear Discriminant Analysis (LDA) algorithm. These models were developed using reflectance data at 940 nm and X-ray classes, individually or in combination (Fig. 7). All models showed good performance in the validation set, with predictive accuracy of 0.96, 0.98 and 0.98 using reflectance, X-ray and the combination of reflectance + X-ray data, respectively.

## Discussion

In seed industry, the development of non-destructive methods for quality control and screening is still a challenge. High-quality seeds are relevant for the entire seed business, which include breeders, producers, traders, variety registration agencies and distributors. Germination and vigor tests are the methods most widely employed for seed quality assessment [17], and the greater seed germination and vigor the better seedling performance in the field. Germination tests determine the ability to produce normal seedlings

under favorable field condition, and vigor tests estimate the potential for rapid and uniform emergence of normal seedlings under a wide range of field conditions [17].

In this study, we present an approach based on two advanced imaging techniques for assessing seed quality using multispectral and X-ray data. Our studies included the combination of these techniques because multispectral imaging has a focus on seed surface profile, which may be negatively affected without reaching important internal regions of seeds. An oilseed plant was used as a model. Initially, seed performance of three lots was tested using conventional analytical methods, which showed that the three seedlots significantly differed in terms of germination capacity and vigor. Results also showed differences in crude fat content, but not for crude protein. A PCA method was applied to the multispectral data, which revealed the most meaningful wavelengths to distinguish the seedlots (780, 850, 880, 940 and 970 nm). The group identified as high-vigor seeds (Lot 2) had the highest crude fat content and the lowest reflectance spectra for both ventral and dorsal seed surfaces. In addition, this group presented more radiographic images in class 1, i.e., with tissues completely filling the seed or slight empty space between the endosperm and the seed coat, which was positively correlated with normal seedlings in the germination test.

There are several reasons to explain why high-vigor seeds have lower reflectance values. When light strikes an object, its rays can be reflected, transmitted, scattered, or absorbed. Reflectance properties of an object depend on its physical and chemical states [18]. In the NIR region, distinctive spectral data correspond to energy absorption of functional groups containing a hydrogen atom (combination of C-H, N-H and O-H) [19]. The 970 nm wavelength is associated with water [20], and there is evidence of absorption peak of fatty tissues at 890 and 940 nm [20, 21]. As a consequence of this, fatty tissues are less reflective as shown in seeds of Lot 2; however, the reflectance data also depends on the color in which the brightest regions are most reflective [22]. For example, there were lighter colored tissues for the ventral and dorsal seed surface in Lot 3 (medium vigor) (Fig. 8) that may have increased the reflectance intensity, i.e., although Lot 3 showed a slight tendency towards higher crude fat content than Lot 1 (low vigor) (Fig. 1), Lot 3 exhibited the highest reflectance (Fig. 2).

High fat content in seeds has been associated with rapid availability of energy and increased mobilization of reserves during germination [1, 23]; therefore, the high percentage of crude fat in Lot 2 may explain the rapid emergence of its seedlings. In previous studies, fat was the main molecule mobilized during germination of *Cereus jamacaru* seeds [24]. Hence, seed oil content is an important seed quality parameter in oilseed species [7].

It has been demonstrated that unhealthy tissues or non-viable seeds are not good absorbers of NIR energy [25 – 28]. This was also verified in Lots 1 and 3, which showed a higher proportion of deteriorated tissues in the radiographic images (seeds in classes 2 and 3), higher reflectance values in the NIR region and more dead seeds in the germination test. A reflectance trend toward higher values in the NIR spectra was also found in non-viable seeds of *Brassica oleracea* [29] and *Ricinus cummunis* [26]. Healthier tissues are metabolically more active; therefore, they absorb more light energy and reflect less.

There was a direct relationship between jatropha seeds with deteriorated tissues in radiographic images and abnormal seedlings or dead seeds in the germination test. But, dead seeds occurred mainly when deteriorated tissues were reaching the embryonic axis (X-ray images in class 3). Lower grayscale in the X-ray images are strongly related to lower physical integrity and less stored reserves, including protein, carbohydrates and fats [30].

All lots used in this study showed a mixture of normal seedlings, abnormal seedlings and dead seeds, therefore, although Lots 1 and 2 had the most extreme quality difference, these lots showed some seeds with similar spectral signature that overlapped in Can 1 (Fig. 4). A similar behavior was also found in the separation of early germinated seeds from medium germination and dead seeds in cowpea using an LDA model [31].

Results obtained in this study would not be achieved using RGB images because they are limited to the visible light spectrum. Moreover, other traditional reflectance techniques such as NIR spectroscopy measure only a relatively small area of a specimen (spot measurement where the sensor is located), so they do not provide spatial information that is important for many seed inspection applications. NIR spectroscopy can be successfully used when the attribute measured related to seed quality (e.g. disease, damage and phenolic compound) is located in specific regions of seeds [11, 32], but this method requires spectral pre-processing methods to remove irrelevant information and improve the performance of calibration models [11, 28, 33, 34]. On the other hand, NIR spectral imaging has provided spatial and spectral information of samples using different wavelengths to obtain rapid and accurate measurements of uniform and non-homogeneous samples. It has been used to predict seed health status [35], discrimination of seeds at different maturation stages [36] or with mechanical damage [37].

Different from multispectral imaging that generates reflectance images, X-ray technique produces transmittance images resulting from short wavelengths that can penetrate seed tissues [38, 39]. X-ray tests are well known in detecting internal seed defects, which contributes to predict problems in the future such as physiological abnormalities during imbibition, germination and seedling development [8, 9, 40, 41]. It has been successfully used to identify insect damages in *Triticum aestivum* [40, 42] and *Glycine max* seeds [43], desiccation sensitivity of *Quercus rubra* seeds [44], and mechanical damage in *Archontophoenix cunninghamii* seeds [45].

Our findings indicate that multispectral and X-ray imaging techniques have the potential to objectively predict seed quality, with high accuracy (0.96–0.98). Jatropha seed is known for its short-term storage [46], therefore, rapid, simple and accurate methods can help producers and distributors to ensure seed quality. However, in-depth studies with a larger number of seedlots from different regions and crop-year are still needed to strengthen the methodologies for applying multispectral and X-ray imaging in the quality inspection of jatropha seeds or other important oilseeds.

## Conclusion

Seed quality is mainly monitored by destructive, laborious and time-consuming methods that require specialized analysts. In this research, we demonstrated two non-destructive techniques for seed quality characterization based on multispectral and X-ray images. Three different classes of jatropha seeds were tested. We proposed an approach using X-ray images to investigate internal aspects of seeds, such as deteriorated tissues in the embryonic axis and endosperm, considering the fact that seed surface can be negatively affected without reaching important internal regions of seeds. We compared multispectral and X-ray data with analytical methods traditionally used to evaluate seed performance, including germination tests, electrical conductivity and seedling emergence. Our results demonstrated that multispectral and X-ray images have a strong relationship with seed physiological potential. Reflectance data at 940 nm and X-ray data, individually or combined, showed above 0.96 accuracy to predict quality traits such as normal seedlings, abnormal seedlings and dead seeds. Multispectral and X-ray imaging could be used for rapid, sustainable and non-destructive evaluation of seed quality in the future, overcoming intrinsic subjectivities of seed testing.

## **Methods**

### **Seed material**

Jatropha plants have a great variation in the fruit ripening time, with the same plant showing fruits at different stages of ripeness [47]. In this study, changes of pericarp color were used as indicators of ripening, and all fruits were collected in the 'brown dry' maturity stage [48]. Three different seedlots (Lots 1, 2 and 3) were investigated. After the harvest, fruits were kept at room temperature for one week. Then, seeds were extracted manually from the fruits and each seedlot was homogenized and evaluated for moisture content (fresh weight basis) which ranged from 11.3 to 11.8%. All seedlots were packed in Kraft paper bags and stored at 20 °C and 40% RH during the experimental period. In this condition, the seed water content was reduced, varying between 6.5 and 6.6%. Traditional tests were performed to rank the lots based on germination and vigor.

### **Traditional tests to rank lots based on germination and vigor**

#### ***Germination tests***

Seeds were sown on paper towel and sand substrates and kept at 30 °C and a photoperiod of 12 hours: ten repetitions of 10 seeds per lot were distributed on paper towels moistened with distilled water (1: 2.5, g: ml), and four replications of 25 seeds per lot were sown in sand (moistened to 60% of its water holding capacity) in plastic trays. The percentage of normal seedlings per lot were recorded at 5 and 10 days after sowing. To calculate the germination rate index – GRI [49], the number of emerged seedlings on paper substrate was monitored daily during 10 days.

#### ***Electrical conductivity***

Four replications of 15 seeds per lot were weighed and maintained for 6 hours in containers with 75 mL of distilled water at 25 °C [50]. The electrical conductivity ( $\mu\text{S cm}^{-1} \text{g}^{-1}$ ) was measured using a DIGIMED DM-32 conductivity meter.

### ***Seedling emergence***

Four subsamples of 25 seeds per lot were sown in plastic trays containing sand moistened to 60% of its water holding capacity. Boxes were maintained at room temperature. The percentage of emerged seedlings was determined at 10 days after sowing.

Data from germination tests, electrical conductivity and seedling emergence were analyzed separately by analysis of variance in a completely randomized design and the means compared by the Tukey's test ( $P < 0.05$ ).

### ***Fat and protein content***

Proximate chemical composition analysis of the seeds was performed according to the methods of the Association of Official Analytical Chemists [51] for crude fat (AOAC N<sup>o</sup>.4.5.01) and crude protein content (AOAC N<sup>o</sup>.4.2.11). Percent data of crude fat and crude protein content were separately fitted to a linear model with normal distribution for errors, including seedlot as the fixed effect in the linear predictor. Post-hoc contrasts between seedlots were further determined by Tukey test ( $P < 0.05$ ).

### **Multispectral imaging**

Multispectral images were obtained using a VideometerLab4 (Videometer A/S, Herlev, Denmark) and its software VideometerLab version 3.14.9. This instrument is integrated with a sphere providing homogeneous and diffuse illumination using strobe light-emitting diode (LED) technology. Reflectance images were captured at 19 different wavelengths (365, 405, 430, 450, 470, 490, 515, 540, 570, 590, 630, 645, 660, 690, 780, 850, 880, 940 and 970 nm), combining them into high-resolution multispectral images (40  $\mu\text{m}/\text{pixel}$ ). Every pixel in the image contains reflectance data, which varies depending on color, texture and chemical composition of the sample.

Ten replications of 10 seeds per lot were placed in 9-cm Petri dishes. Before image acquisition, the individual and automated adjustment of light intensity in each wavelength band was performed to optimize the illumination for the specific type of sample, resulting in an improved signal-to-noise ratio in such a way that the multispectral images captured from different seed classes could be directly comparable. Light setup was adjusted using a representative sample area, then the strobe time of each illumination type was optimized with respect to this area. The auto light assures an optical dynamic range of each band without saturation within the auto light ROI. Subsequently, the instrument was calibrated using three calibration targets: (i) uniform bright disc, (ii) uniform dark disc, and (iii) geometric disc which is black with dots in a rectangular grid.

Multispectral images were captured from both ventral and dorsal seed surface of 10 samples with 10 seeds per lot. The overview of the ventral and dorsal surfaces of the three seedlots is shown in Fig. 8. After successive lighting using 19 LEDs (sequential strobes), multispectral images of a sample (plate with 10 seeds) were captured in a few seconds, requiring no sample preparation. The ROI of each seed was extracted into a Binary Large Object (BLOB) toolbox, a built-in function in VideometerLab software; each BLOB was a representation of one seed. Mean spectra were plotted to show the difference among the three seedlots based on their multispectral signatures. A normalized canonical discriminant analysis (nCDA) algorithm was used as a supervised model based on multispectral image transformation, which allows to minimize the distance to observations within seedlot and to maximize the distance to observations among seedlots.

We applied a PCA method to process the multispectral data using the “FactoMiner” package [52]. A biplot using the first two principal components (PC1 and PC2) was built to select the most meaningful wavelengths to discriminate the seedlots, according to Pearson’ correlation test ( $P < 0.05$ ). Multispectral data corresponding to only meaningful wavelengths, as previously assigned by PCA were used in a CDA model implemented with a “candisc” package [53]. We tested the effect of low, high and medium vigor (i.e., three classes of seed physiological potential) on the multispectral data using a multivariate analysis of variance (MANOVA). The statistical analyses were performed using VideometerLab software and the “free software environment for statistical computing and graphics” R [54].

## **X-ray imaging**

In total, 100 seeds per lot were radiographed. Seeds were numbered and fixed on an adhesive paper in groups of 10 seeds. Radiographic images were generated using a MultiFocus digital radiography system (Faxitron Bioptics LLC, USA). This system is equipped with a complementary metal-oxide-semiconductor (CMOS) X-ray sensor coupled with an 11  $\mu\text{m}$  focal spot tube and up to 8X geometric magnification and provides as high as 6  $\mu\text{m}$  resolution for seed imaging with a choice of a 48  $\mu\text{m}$  or 24  $\mu\text{m}$ . The built-in advanced Automatic Exposure Control selects the appropriate exposure time and kV settings for each sample.

After X-ray imaging, four repetitions of 25 seeds were sown in sand (moistened to 60% of its water holding capacity) placed in plastic boxes (32.0 x 28.0 x 10.0 cm), kept at 30 °C and photoperiod of 12 hours. At 10 days after sowing, the individual seeds were evaluated for different quality traits: normal seedlings, abnormal seedlings and dead seeds. Next, they were separated into three different classes based on seed performance in the germination test and tissue integrity in the radiographic images. A CDA analysis was implemented by “candisc” package in R [53] to provide the best discrimination among seedlots categories using a dataset derived from X-ray classes, reflectance data at 940 nm and quality traits (normal seedling, abnormal seedling and dead seed).

## **Validation**

Three models were developed using LDA algorithm. The first model was created using multispectral data at 940 nm. Data obtained from X-ray classes were used to develop the second model. Finally, multispectral and X-ray data were combined to create the third classification model. In total, 300 seeds were used to develop the models. Training was run with 210 seeds (70%), and the remaining 90 seeds (30%) were used for independent validation set. Additionally, 5-fold cross-validation was performed using training data. The metrics of accuracy, Cohen's Kappa coefficient, sensitivity and specificity were calculated using a confusion matrix to evaluate the models. Data analysis was performed by R software using the “caret” package [55].

## **Abbreviations**

ANOVA: Analysis of variance; BLOB: Binary large object; CDA: Canonical discriminant analysis; CMOS: Complementary metal-oxide-semiconductor; GRI: Germination rate index; LDA: Linear discriminant analysis; LED: Light-emitting diode; MANOVA: Multivariate analysis of variance; nCDA: Normalized canonical discriminant analysis; NIR: Near infrared; PCA: Principal component analysis; ROIs: Region of interest; RU: Relative humidity.

## **Declarations**

### **Acknowledgments**

The authors gratefully acknowledge financial support from the Sao Paulo Research Foundation–FAPESP.

### **Authors' contributions**

CBS generated research ideas. VJB and LCASS conducted the experiments. CBS and VJB wrote the first draft. GMM and VJB performed statistical data analyses. CBS edited the manuscript. CBS, VA, JMC and BB reviewed the manuscript for technical and scientific accuracy. CBS wrote the final manuscript, funding acquisition, supervised the project. All authors read and approved the final manuscript.

### **Funding**

This work was supported by Sao Paulo Research Foundation–FAPESP (Grant#2017/15220-7, Grant#2018/03802-4, Grant#2018/03807-6, Grant#2018/01774-3 and Grant#2019/04127-1).

### **Availability of data and materials**

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### **Ethics approval and consent to participate**

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Crop Science, College of Agriculture “Luiz de Queiroz”, University of São Paulo, 11 Padua Dias Ave., Box 9, 13418-900, Piracicaba, SP, Brazil.

<sup>2</sup>Laboratory of Environmental Microbiology, Brazilian Agricultural Research Corporation, Embrapa Environment, Rodovia SP 340, Km 127.5, 13820-000 Jaguariúna, Brazil.

<sup>3</sup>Laboratory of Radiobiology and Environment, Center for Nuclear Energy in Agriculture, University of São Paulo, 303 Centenario Ave., Sao Dimas, 13416-000, Piracicaba, SP, Brazil.

<sup>4</sup>Technical University of Denmark, DK-2800 Lyngby, Denmark

<sup>5</sup>Department of Agroecology, Science and Technology, Aarhus University, Slagelse 4200, Denmark.

## References

1. Finch-Savage WE, Bassel GW. Seed vigour and crop establishment: extending performance beyond adaptation. *J Exp Bot.* 2016;67;567–1. <https://doi.org/10.1093/jxb/erv490>.
2. Atkinson JL, McCarty LB, Yelverton F, McElroy S, Bridges, WC. Doveweed (*Murdannia nudiflora*) response to environmental resource availability and cultural practices. *Weed Sci.* 2019;67;214–0. <https://doi.org/10.1017/wsc.2018.89>.
3. Huang M, Wang QG, Zhu QB, Qin JW, Huang G. Review of seed quality and safety tests using optical sensing technologies. *Seed Sc Technol.* 2015;43;337–6. <https://doi.org/10.15258/sst.2015.43.3.16>.
4. Manley M, Williams P, Nilsson D, Geladi P. Near infrared hyperspectral imaging for the evaluation of endosperm texture in whole yellow maize (*Zea mays* L) kernels. *J Agric Food Chem.* 2009;57;8761–9. <https://doi.org/10.1021/jf9018323>.
5. Singh CB, Jayas DS, Paliwal J, White NDG. Detection of midge-damaged wheat kernels using short-wave near-infrared hyperspectral and digital colour imaging. *Biosyst Eng.* 2010;105;380–7. <https://doi.org/10.1016/j.biosystemseng.2009.12.009>.
6. França-Silva F, Rego CHQ, Gomes-Junior FG, Moraes MHD, Medeiros AD, Barboza da Silva C. Detection of *Drechslera avenae* (Eidam) Sharif [*Helminthosporium avenae* (Eidam)] in black oat seeds (*Avena strigosa* Schreb) using multispectral imaging. *Sensors* 2020;20;1–0. <https://doi.org/10.3390/s20123343>.

7. Rahman A, Cho BK. Assessment of seed quality using non-destructive measurement techniques: a review. *Seed Sci Res.* 2016;26;285– <https://doi.org/10.1017/S0960258516000234>.
8. Arkhipov MV, Priyatkin NS, Guskova LP, Potrakhov NN, Gryaznov AY, Bessonov VB, Obodovskii AV, Staroverov NE. X-ray computer methods for studying the structural integrity of seeds and their importance in modern seed science. *Tech Phys.* 2019;64;582–2. <https://doi.org/10.1134/S1063784219040030>.
9. Kotwaliwale N, Singh K, Kalne A, Jha SN, Seth N, Kar A. X-ray imaging methods for internal quality evaluation of agricultural produce. *J Food Sci. Technol.* 2014;51;1–5. <https://doi.org/10.1007/s13197-011-0485-y>.
10. ElMasry G, Mandour N, Al-Rejaie S, Belin E, Rousseau D. Recent applications of multispectral imaging in seed phenotyping and quality monitoring— an overview. *Sensors* 2019;19;1–2. <https://doi.org/10.3390/s19051090>.
11. Medeiros AD, Silva LJ, Ribeiro JPO, Ferreira KC, Rosas JTF, Santos AA, Barboza da Silva, C. Machine learning for seed quality classification: An advanced approach using merger data from FT-NIR spectroscopy and X-ray imaging. *Sensors* 2020;20;1–2. <https://doi.org/10.3390/s20154319>.
12. Mukherjee P, Varshney A, Johnson T, Jha T. *Jatropha curcas*: a review on biotechnological status and challenges. *Plant Biotechnol Rep.* 2011;5;197–5. <https://doi.org/10.1007/s11816-011-0175-2>.
13. Makkar H, Kumar V, Oyeleye OO, Akinleye AO, Angulo-Escalante MA, Becker K. *Jatropha platyphylla*, a new non-toxic jatropha species: physical properties and chemical constituents including toxic and antinutritional factors of seeds. *Food Chem.* 2011;125;63–1. <https://doi.org/10.1016/j.foodchem.2010.08.037>.
14. Raheman H, Mondal S. Biogas production potential of jatropha seed cake. *Biomass Bioenerg.* 2012;37;25–0. <https://doi.org/10.1016/j.biombioe.2011.12.042>.
15. Mulpuri S, Carels N, Bahadur B. *Jatropha*, challenges for a new energy crop. third ed. A sustainable multipurpose crop. Springer: Singapore. 2019.
16. Steinbrück J, Tavakkol S, Francis G, Bockhorn H. *Jatropha* – Potential of biomass steam processing to convert crop residues to bio-coal and thus triple the marketable energy output per unit plantation area. *Ind Crop Prod.* 2019;136;59–5. <https://doi.org/10.1016/j.indcrop.2019.04.065>.
17. Baalbaki R, Elias S, Marcos-Filho J, McDonald M. (Ed.). Seed vigor testing handbook. Contribution to the Handbook on seed testing 32. Association of Official Seed Analysts: Ithaca. 2009.
18. Elmasry G, Sun DW. Principles of hyperspectral imaging technology. In: *Hyperspectral imaging for food quality analysis and control*. London: Elsevier; 2010. p. 3-43. <https://doi.org/10.1016/B978-0-12-374753-2.10001-2>.
19. Osborne BG, Fearn T, Hindle PH. *Practical NIR spectroscopy with applications in food and beverage analysis*. Longman Scientific and Technical: Harlow. 1993.
20. Sendin K, Manley M, Williams PJ. Classification of white maize defects with multispectral imaging. *Food Chem.* 2018;243;311–8. <https://doi.org/10.1016/j.foodchem.2017.09.133>.
21. Jue T, Masuda K. *Application of near infrared spectroscopy in biomedicine*. Springer, New York. 2013.

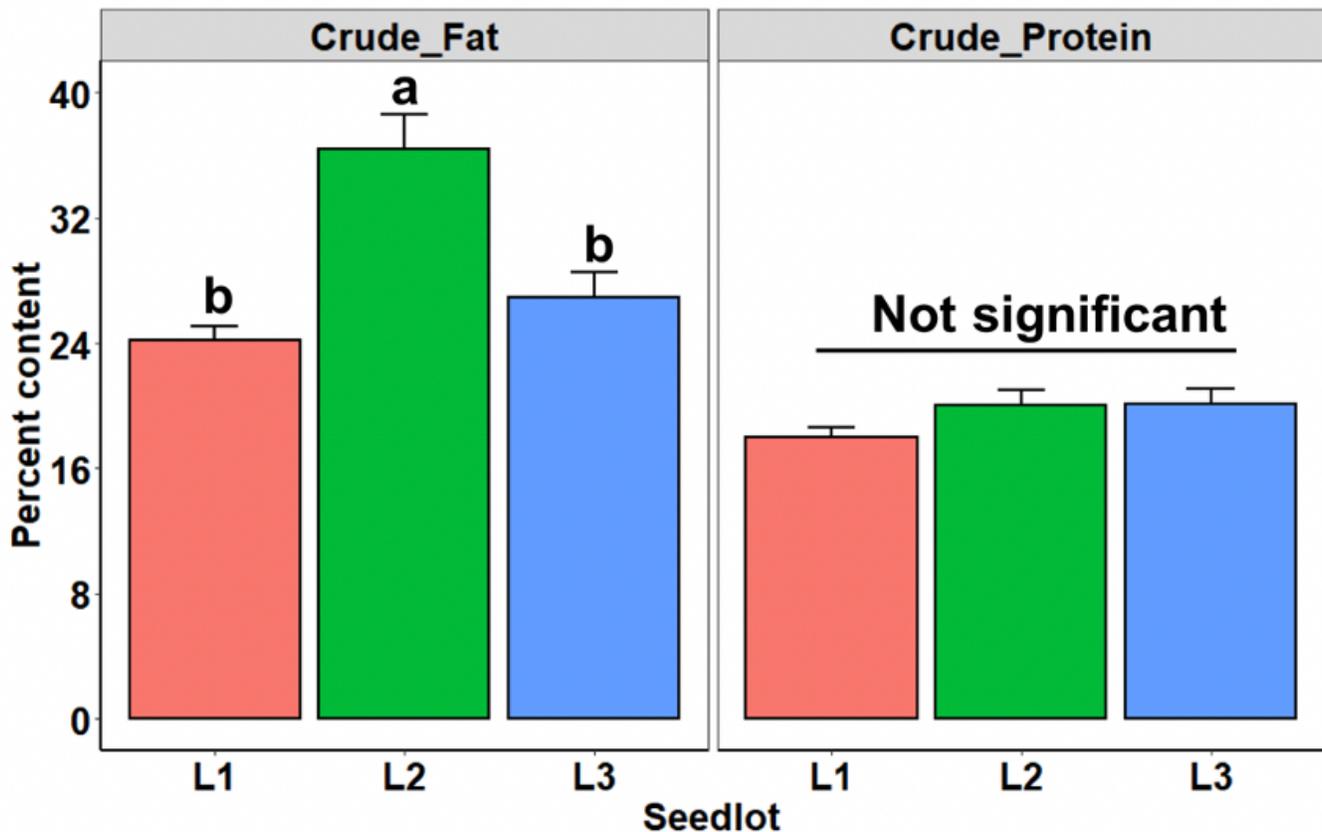
22. Zhang X, Liu F, He Y, Li X. Application of hyperspectral imaging and chemometric calibrations for variety discrimination of maize seeds. *Sensors* 2012;12;17234–6.  
<https://doi.org/10.3390/s121217234>.
23. Kelly AA, Quettier AL, Shaw E, Eastmond PJ. Seed storage oil mobilization is important but not essential for germination or seedling establishment in *Arabidopsis*. *Plant Physiol.* 2011;157;866–5.  
<https://doi.org/10.1104/pp.111.181784>.
24. Alencar NLM, Innecco R, Gomes-Filho E, Gallao MI, Alvarez-Pizarro JC, Prisco JT, Oliveira AB. Seed reserve composition and mobilization during germination and early seedling establishment of *Cereus jamacaru* C. ssp. *jamacaru* (Cactaceae). *An Acad Bras Cienc.* 2012;84;823–2.  
<https://doi.org/10.1590/S0001-37652012000300024>.
25. Dumont J, Hirvonen T, Heikkinen V, Mistretta M, Granlund L, Himanen K, Fauch L, Porali I, Hiltunen J, Keski-Saari S, Nygren M, Oksanen E, Hauta-Kasari M, Keinänen M. Thermal and hyperspectral imaging for Norway spruce (*Picea abies*) seeds screening. *Comput Electron Agr.* 2015;116;118–4.  
<https://doi.org/10.1016/j.compag.2015.06.010>.
26. Olesen MHR, Nikneshan P, Shrestha S, Tadayyon A, Deleuran LC, Boelt B, Gislum R. Viability prediction of *Ricinus communis* seeds using multispectral imaging. *Sensors* 2015;15;4592–4.  
<https://doi.org/10.3390/s150204592>.
27. Kandpal LM, Lohumi S, Kim MS, Kang JS, Cho BK. Near-infrared hyperspectral imaging system coupled with multivariate methods to predict viability and vigor in muskmelon seeds. *Sensor Actuator.* 2016;229;534–4. <https://doi.org/10.1016/j.snb.2016.02.015>.
28. Xia Y, Xu Y, Li J, Zhang C, Fan S. Recent advances in emerging techniques for non-destructive detection of seed viability: A review. *Artif Intell Agr.* 2019;01;35–7.  
<https://doi.org/10.1016/j.aiia.2019.05.001>.
29. Shetty N, Min T-G, Gislum R, Olesen MH, Boelt B. Optimal sample size for predicting viability of cabbage and radish seeds based on near infrared spectra of single seeds. *J Near Infrared Spectrosc.* 2011;19; 451–1. <https://doi.org/10.1255/jnirs.966>.
30. Medeiros AD, Pinheiro DT, Xavier WA, Silva LJ, Dias DCFS. Quality classification of *Jatropha curcas* seeds using radiographic images and machine learning. *Ind Crop Prod.* 2020;146;01–7.  
<https://doi.org/10.1016/j.indcrop.2020.112162>.
31. Elmasry G, Mandour N, Wagner MH, Demilly D, Verdier J, Belin E, Rousseau D. Utilization of computer vision and multispectral imaging techniques for classification of cowpea (*Vigna unguiculata*) seeds. *Plant Methods* 2019;15;1-6. <https://doi.org/10.1186/s13007-019-0411-2>.
32. Wang D, Dowell FE, Ram MS, Schapaugh WT. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *Int J Food Prop.* 2004;7;75–2, <https://doi.org/10.1081/JFP-120022981>.
33. Vrešak M, Olesen HM, Gislum R, Bavec F, Jørgensen, JR. The use of image-spectroscopy technology as a diagnostic method for seed health testing and variety identification. *PLOS ONE* 2016;11;1–0.  
<https://doi.org/10.1371/journal.pone.0152011>.

34. Levasseur-Garcia, C. Updated overview of infrared spectroscopy methods for detecting mycotoxins on cereals (corn, wheat, and barley). *Toxins* 2018;10;1–3. <https://doi.org/10.3390/toxins10010038>.
35. Qiao X, Jiang J, Qi X, Guo H, Yuan D. Utilization of spectral-spatial characteristics in shortwave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food Chem.* 2017;220;393–9. <https://doi.org/10.1016/j.foodchem.2016.09.119>.
36. Rodríguez-Pulido FJ, Barbin DF, Sun D-W, Gordillo B, González-Miret ML, Heredia FJ. Grape seed characterization by NIR hyperspectral imaging. *Postharvest Biol Technol.* 2013;76:74–2. <https://doi.org/10.1016/j.postharvbio.2012.09.007>.
37. Salimi Z, Boelt B. Classification of processing damage in sugar beet (*Beta vulgaris*) seeds by multispectral image analysis. *Sensors* 2019;19;1–9. <https://doi.org/10.3390/s19102360>.
38. Neethirajan S, Jayas DS, White NDG. Detection of sprouted wheat kernel using soft X-ray image analysis. *J Food Eng.* 2007;81;509–3. <https://doi.org/10.1016/j.jfoodeng.2006.11.020>.
39. Kotwaliwale N, Singh K, Kalne A, Jha SN, Seth, N, Kar A. X-ray imaging methods for internal quality evaluation of agricultural produce. *J Food Sci Technol.* 2014;51;1–5. <https://doi.org/10.1007/s13197-011-0485-y>.
40. Karunakaran C, Jayas DS, White NDG. Identification of wheat kernels damaged by the red flour beetle using X-ray images. *Biosyst Eng.* 2004;87;267–4. <https://doi.org/10.1016/j.biosystemseng.2003.12.002>.
41. Mathanker SK, Weckler PR, Bowser TJ. X-ray applications in food and agriculture: a review. *T ASABE* 2013;56;1227–9. <https://doi.org/10.13031/trans.56.9785>.
42. Karunakaran C, Jayas DS, White NDG. Detection of internal wheat seed infestation by *Rhizopertha dominica* using X-ray imaging. *J Stored Prod Res.* 2004;40;507–6. <https://doi.org/10.1016/j.jspr.2003.06.003>.
43. Chelladurai V, Karupppiah K, Jayas DS, Fields PG, White NDG. Detection of *Callosobruchus maculatus* (F.) infestation in soybean using soft X-ray and NIR hyperspectral imaging techniques. *J Stored Prod Res.* 2014;57;43– <https://doi.org/10.1016/j.jspr.2013.12.005>.
44. Goodman RC, Jacobs DF, Karrfalt RP. Evaluating desiccation sensitivity of *Quercus rubra* acorns using X-ray image analysis. *Can J For Res.* 2005;35;2823–1. <https://doi.org/10.1139/x05-209>.
45. Sturião WP, Landgraf PRC, Rosa TP. X-ray test for the evaluation of the internal morphology of the seeds of seafortia palm trees. *Semin-Cienc Agrar.* 2012;33;3003–8. <https://doi.org/10.5433/1679-0359.2012v33Supl1p3003>.
46. Moncaleano-Escandon J, Silva BCF, Silva SRS, Granja JAA, Alves MCJL, Pompelli MF. Germination responses of *Jatropha curcas* L. seeds to storage and aging. *Ind Crop Prod.* 2013;44;684–0. <https://doi.org/10.1016/j.indcrop.2012.08.035>.
47. Silip JJ. Lifecycle duration and maturity heterogeneity of *Jatropha curcas* *J Sustain Develop.* 2010;3;291-5. <https://doi.org/10.5539/jsd.v3n2p291>.
48. Brito CD, Loureiro MB, Souza-Junior AP, Fernandez LG, Castro RD. Morphophysiological profile of *Jatropha curcas* fruits and seeds maturation. *Semin-Cienc Agrar* 2015;36;3615-8.

<https://doi.org/10.5433/1679-0359.2015v36n6p3615>.

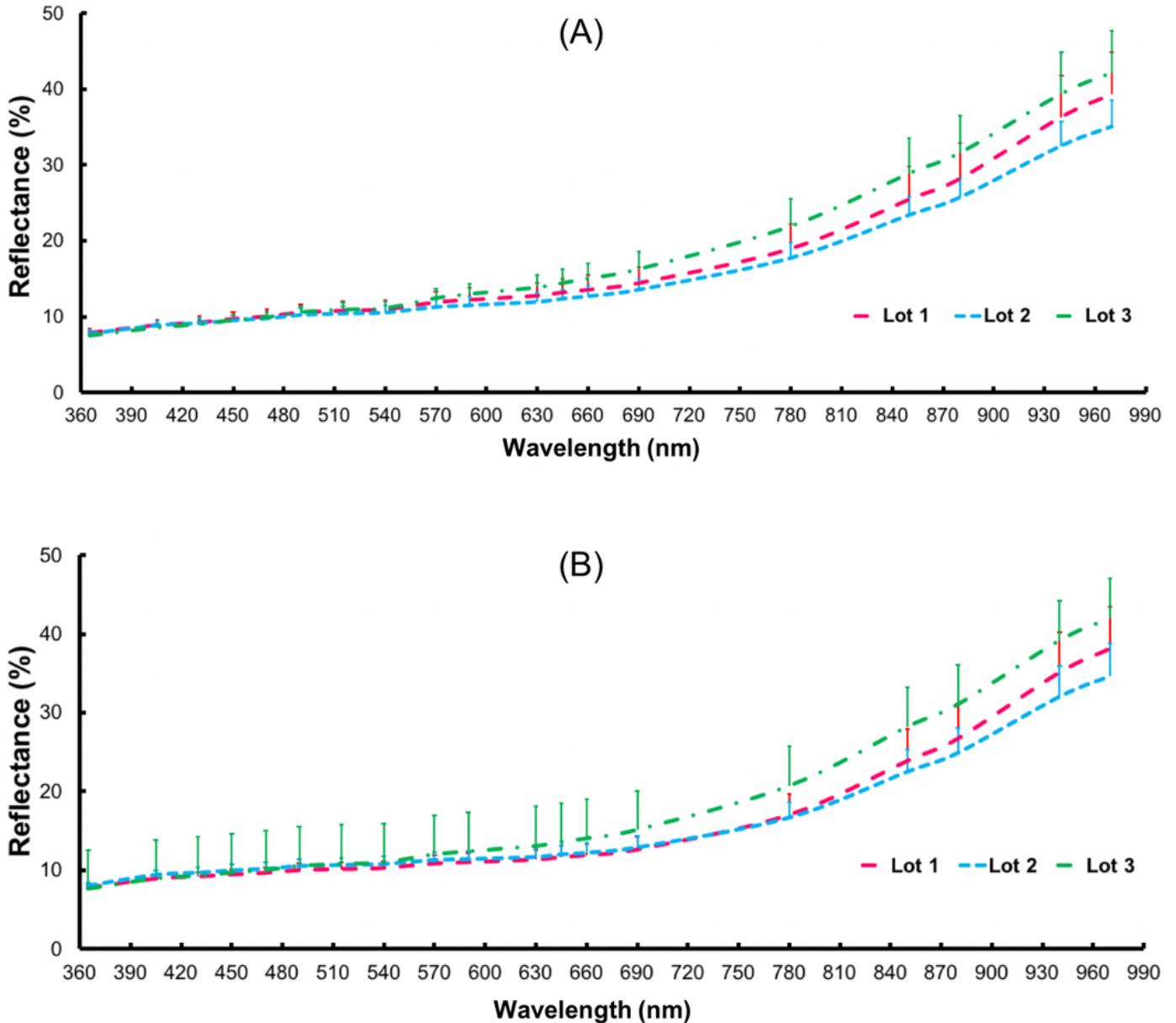
49. Maguire JD. Speed of germination-aid in selection and evaluation for seedling emergence and vigor. *Crop Sci.* 1962;2;176–7. <http://dx.doi.org/10.2135/cropsci1962.0011183X000200020033x>.
50. Araujo RF, Araujo EF, Donzeles SML, Costa GM, Zonta JB. Electrical conductivity test for physic nut (*Jatropha curcas*). *Idesia* 2011;29;79–6. <http://doi.org/10.4067/S0718-34292011000200010>.
51. Horwitz W, Latimer GW (Ed.). Official methods of analysis of AOAC International. 18<sup>th</sup> AOAC International: Gaithersburg. 2011.
52. Lê S, Josse J, Husson F. FactoMineR: A package for multivariate analysis. *J Stat Softw.* 2008;25;1–8. <https://doi.org/10.18637/jss.v025.i01>.
53. Friendly M, Fox J. candisc: Visualizing Generalized Canonical Discriminant and Canonical Correlation Analysis. R package version 0.8-0. 2017. Retrieved from <https://CRAN.R-project.org/package=candisc>.
54. R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2019. Retrieved from <https://www.r-project.org/>.
55. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* 2008, 28. <https://doi.org/10.18637/jss.v028.i05>.

## Figures



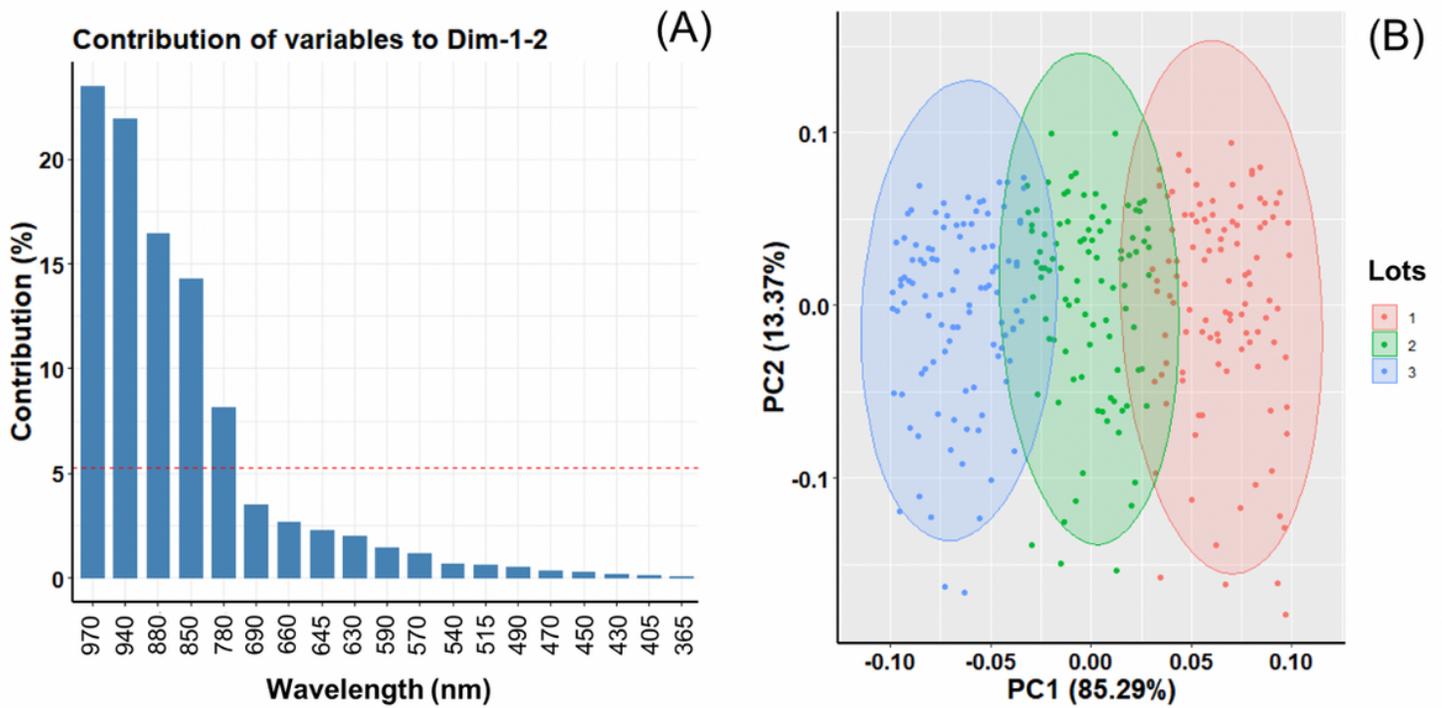
**Figure 1**

Crude fat and crude protein content of three seedlots of *Jatropha curcas* categorized by different vigor status (low vigor: = Lot 1, high vigor = Lot 2, and medium vigor = Lot 3). ANOVA was performed separately for each variable, and different letters indicate significant contrasts between means (bars + SE, upper standard error) according to Tukey HSD test ( $P < 0.05$ ).



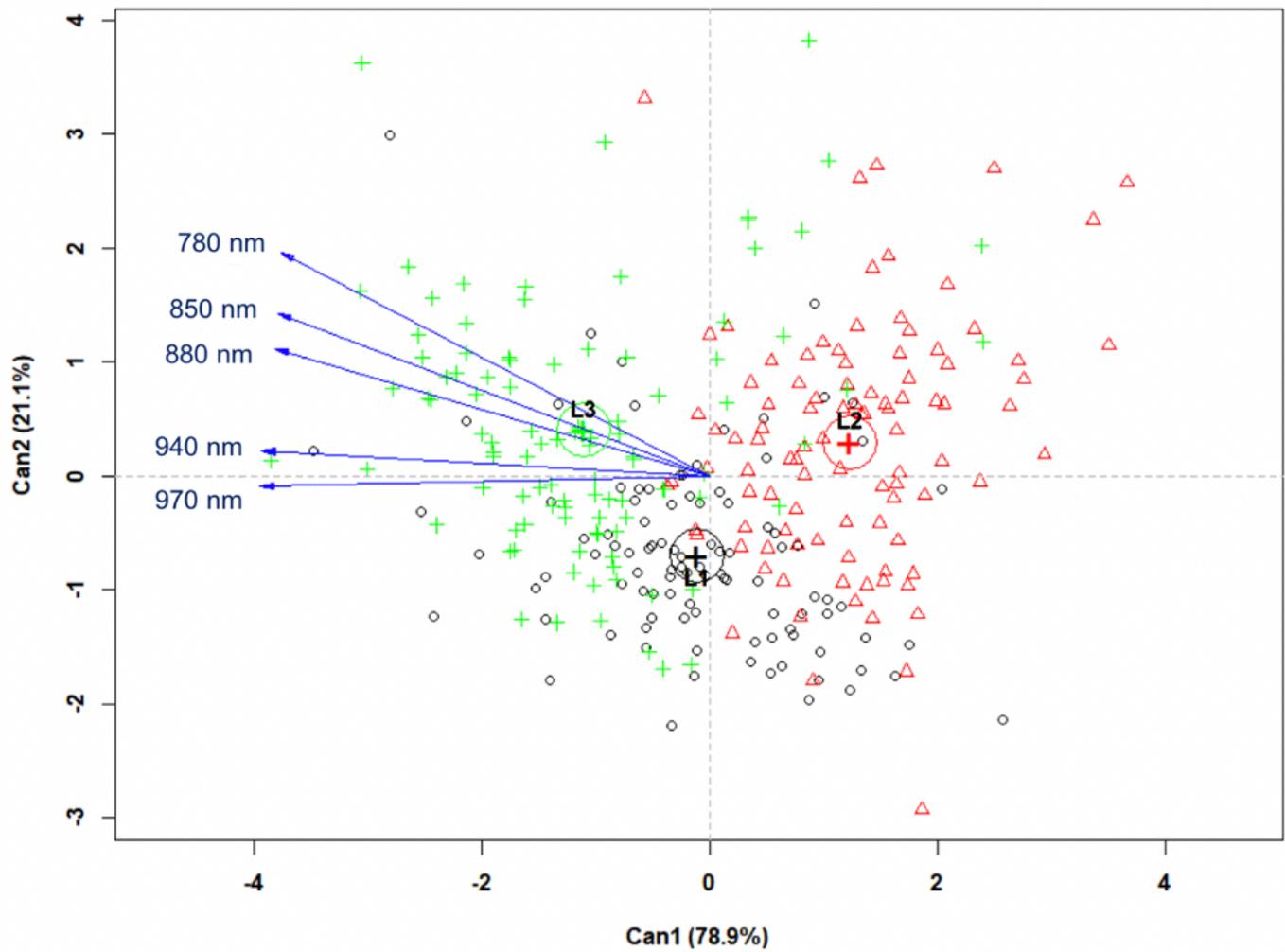
**Figure 2**

Plot of mean spectrum within each region of interest – ROI ( $n = 100$  seeds per lot) at 19 wavelengths for ventral (A) and dorsal (B) seed surface of three lots of *Jatropha curcas* with different vigor levels: low vigor = Lot 1, high vigor = Lot 2, and medium vigor = Lot 3. Vertical bars represent standard deviation (in upper directions).



**Figure 3**

(A) Contributions of response variables to dimensions 1 and 2 with red-dashed line indicating the cut-off for significant variables according to Pearson's correlation test ( $P < 0.05$ ). (B) Biplots of principal component analysis (PCA) for multispectral reflectance in *Jatropha curcas* seedlots (Lots 1, 2 and 3). The five wavelengths from 780 to 970 nm significantly explained the differences among seedlots ( $F = 32.22$ ,  $df = 10, 586$ ,  $P < 0.0001$ ).



**Figure 4**

Canonical discriminant analysis (CDA) of reflectance data obtained at 780, 850, 880, 940 and 970 nm from three seedlots of *Jatropha curcas* with different vigor levels: low vigor = L1 [black circles], high vigor = L2 [red triangles], and medium vigor = L3 [green crosses]) (n = 100 seeds per lot).

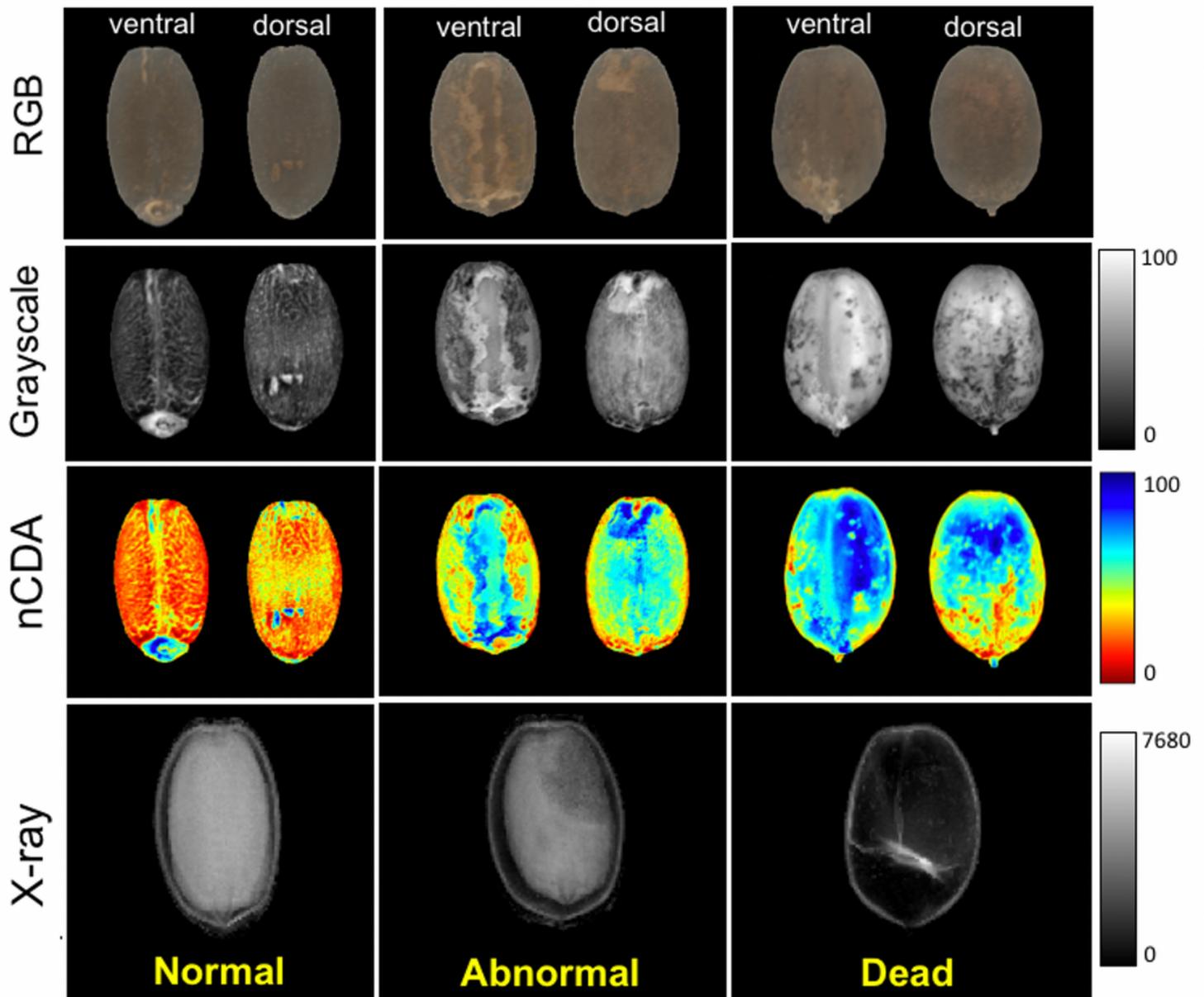
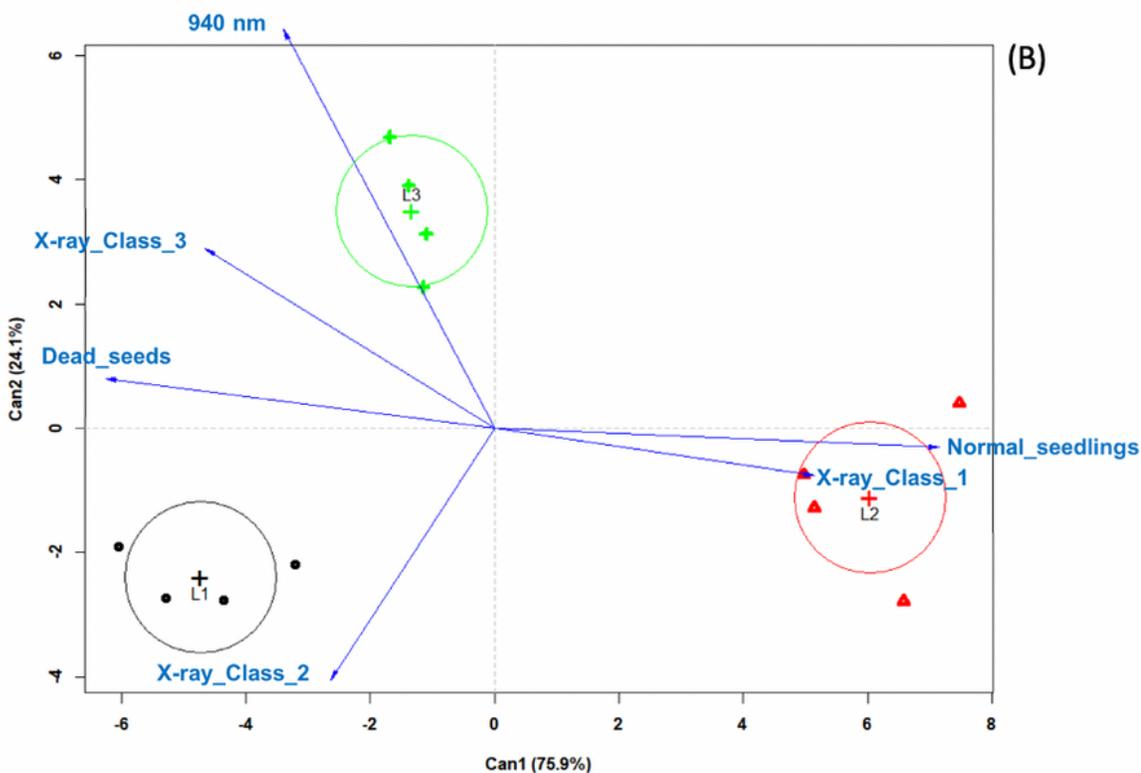
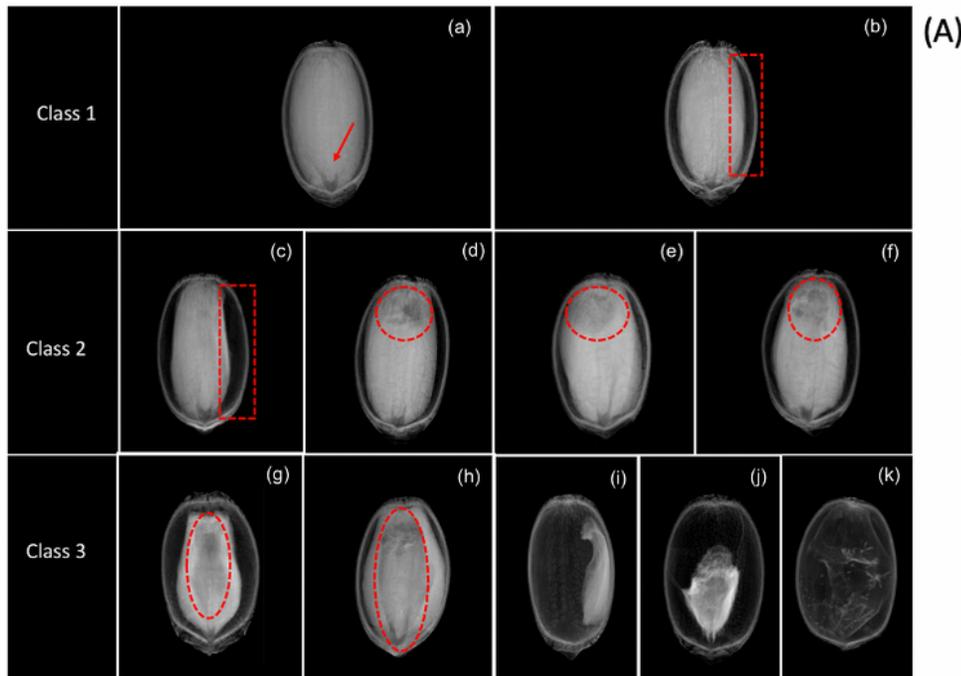


Figure 5

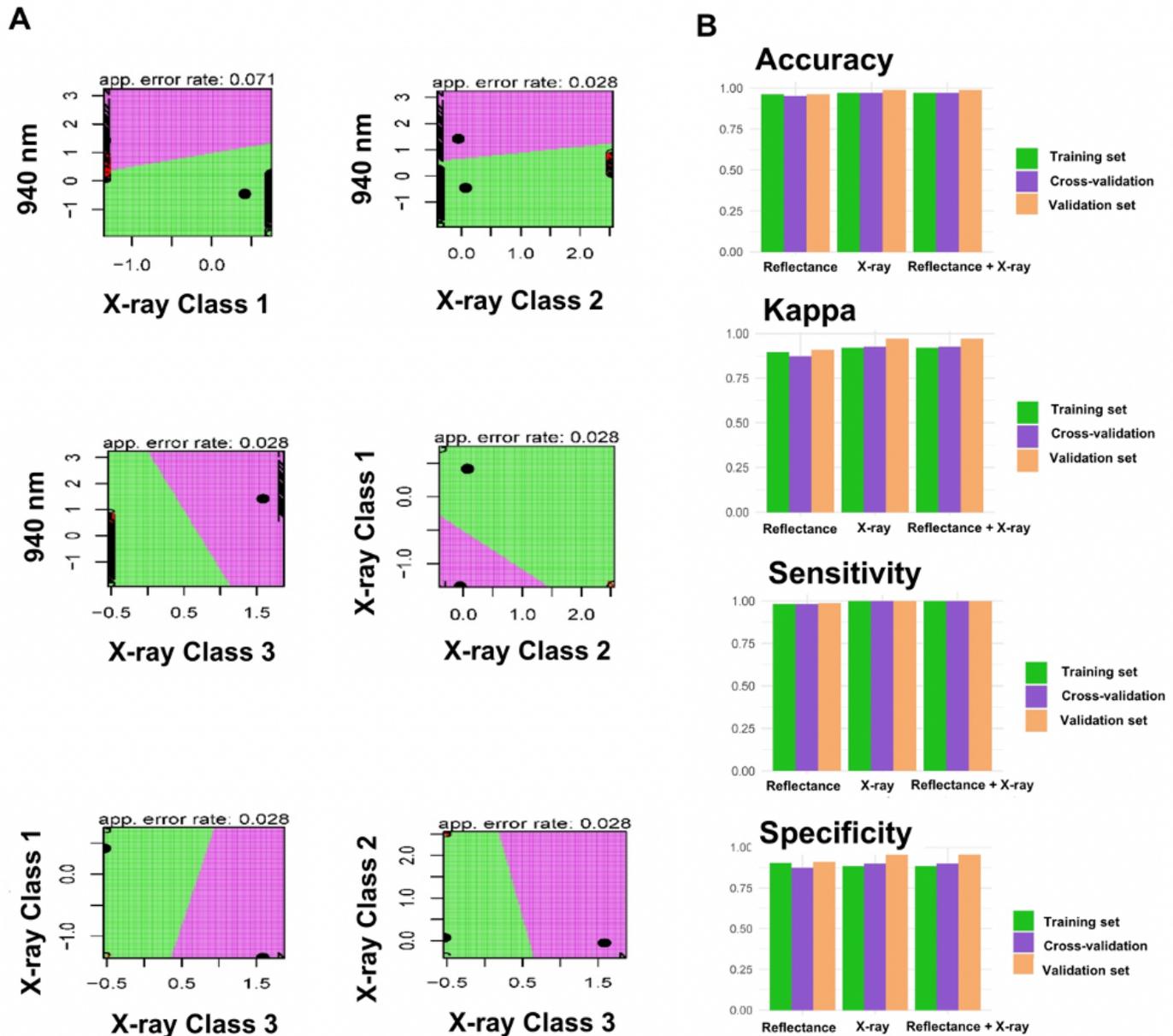
Raw RGB images of ventral and dorsal surfaces of *Jatropha curcas* seeds based on germination capacity and corresponding reflectance images captured at 940 nm (grayscale and transformed images using nCDA algorithm) and X-ray images. RGB images are represented by three-color channels (red, green and blue) to generate a single-color value for each pixel in the image. In the grayscale and transformed images using nCDA each pixel is represented by a single-value that correspond to reflectance intensity. Higher pixel values in the X-ray images indicate higher tissue density.



**Figure 6**

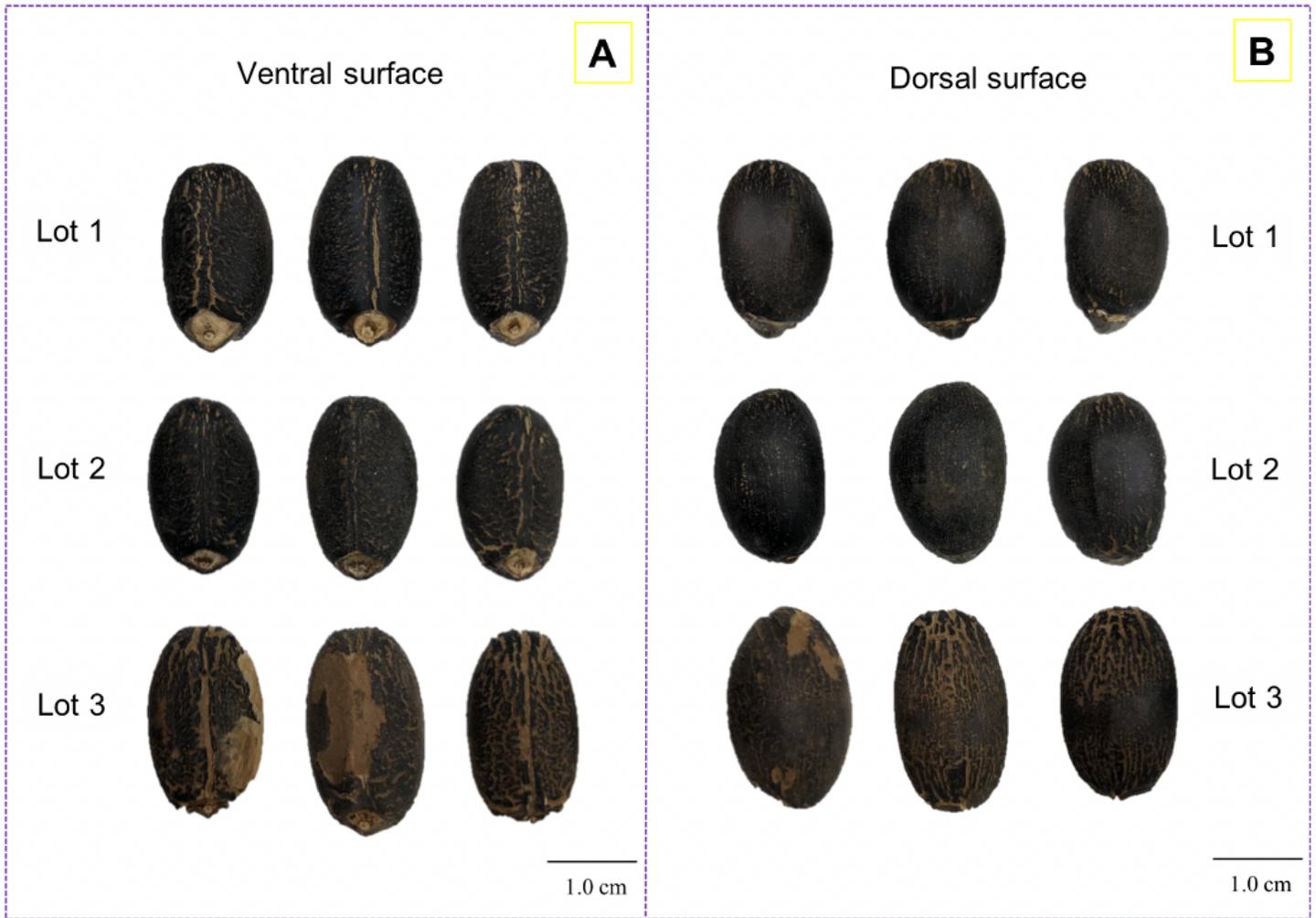
(A) X-ray images of *Jatropha curcas* seeds separated into three different classes based on tissue integrity: Class 1 – (a) tissues completely filling the seed [arrow indicates the embryonic axis region], (b) slight empty spaces ( $\leq 1.23$  mm) between the endosperm and the seed coat; Class 2 – (c) large empty spaces ( $>1.24$  mm) between the endosperm and the seed coat; (d,e,f) deteriorated tissues without reaching the embryonic axis; Class 3 – (g,h) deteriorated tissues reaching the embryonic axis, (i,j)

malformed seeds, (k) empty seed. (B) Canonical discriminant analysis (CDA) of X-ray classes, quality traits (normal seedling and dead seed) and reflectance data at 940 nm for three seedlots with different vigor levels (low vigor = L1 [black circles], high vigor = L2 [red triangles], and medium vigor = L3 [green crosses]); abnormal seedling variable did not appear in the model because it is collinear with dead seed (n = 4 repetitions of 25 seeds).



**Figure 7**

Quality prediction of *Jatropha curcas* seeds based on reflectance at 940 nm and X-ray classes using a Linear Discriminant Analysis (LDA). (A) Plots partitioning two variables and apparent error rates; colored regions delineate each classification area and the observations (spots) within a region is predicted to be from a specific class. (B) Metrics used to validate the models based on accuracy, kappa, sensitivity and specificity.



**Figure 8**

Overview of ventral (A) and dorsal (B) surface of three seedlots (Lots 1, 2 and 3) of *Jatropha curcas*.