INTERNATIONAL
Agrophysics
www.international-agrophysics.org

Int. Agrophys., 2026, 40, 159-172
doi: 10.31545/intagr/216408

Spectral signatures of the physicochemical quality of white, black, red, and parboiled rice processed using non-destructive technologies (VIS/SWIR) combined with machine learning algorithms**

Marina V. Arf[1,2], *Ênio A. Manfroi Filho*[1,2], *Nairiane D.S. Bilhalva*[2], *Dthenifer C. Santana*[1], *Paulo E. Teodoro*[1], *Larissa P.R. Teodoro*[1], *Paulo C. Coradi*[2]*

[1]Laboratory of Digital Agriculture, Campus of Chapadão do Sul, Federal University of Mato Grosso do Sul, Rodovia MS-306, km 105, 79560-000, Chapadão do Sul, MS, Brazil
[2]Laboratory of Postharvest (LAPOS), Campus Cachoeira do Sul, Federal University of Santa Maria, Rua Ernesto Barros, 1345, 96506-322, Cachoeira do Sul, RS, Brazil

Received August 8, 2025; accepted January 5, 2026

A b s t r a c t. This study aimed to correlate the physicochemical attributes of rice grains using non-destructive techniques combined with machine learning algorithms. Samples of white, black, red, and parboiled rice were analyzed using hyperspectral spectroscopy (350-2 500 nm) and subjected to linear regression (LR), random forest (RF), gradient boosting (GB), support vector machine (SVM), convolutional neural networks (CNN), and recurrent neural networks (RNN) prediction. Spectral and physicochemical data were examined through multivariate analysis (PCA) and cross-validation using performance metrics such as R, R^2, MAE, and RMSE. The results indicated that SVM, RF, and GB outperformed the other algorithms, showing higher accuracy and lower variability in the prediction, SVM reached R = 0.952 and R^2 = 0.904, with MAE = 0.409 and RMSE = 0.582, followed closely by RF with R = 0.950 and MAE = 0.416, and GB with R = 0.947 and MAE = 0.431. Black rice stood out for its high protein, lipid, and ash contents. Parboiled rice showed higher fiber content, and white rice was notable for its elevated starch content. Hyperspectral spectroscopy proved effective in differentiating rice types, enabling the identification of relevant spectral bands for optimized sensors. Overall, integrating non-destructive technologies with machine learning shows strong potential for industrial applications.

K e y w o r d s: *Oryza sativa* L., NIR spectroscopy, shallow learning, deep learning, non-destructive methods

*Corresponding author e-mail: paulocoradi@yahoo.com.br
**This work was funded by CAPES (Coordination for the Improvement of Higher Education Personnel) Financial Code 001, CNPq (National Council for Scientific Technological Development) number 304966/2023-1, and FAPERGS-RS (Research Support Foundation of the State of Rio Grande do Sul) number 24/2551-0001150-1.

1. INTRODUCTION

Rice (*Oryza sativa* L.) plays an essential role in human nutrition, serving as a staple food for billions of people across different regions of the world, not only as an energy source but also as a functional food with nutritionally relevant compounds (Brotman *et al.*, 2021). Beyond its nutritional function, this cereal is deeply rooted in cultural, social, and economic contexts, assuming multiple forms of consumption such as white, red, black, and parboiled rice. Each of these variants presents specific characteristics that determine preferences for grain type, viscosity, and aroma, which can vary according to regions and cultures, thereby influencing breeding and marketing strategies (Chen *et al.*, 2021). The diversity among rice types is directly related to differences in the chemical composition, such as anthocyanin and resistant starch content (Pereira *et al.*, 2023).

With the growing demand for nutritious and high-quality foods, there has been an increased search for methods capable of rapidly and non-destructively assessing these properties. In this context, spectroscopy covering visible light and the short-wave infrared range (VIS/SWIR) has gained prominence as an efficient tool for the compositional analysis of grains (Lin *et al.*, 2021). VIS/SWIR spectroscopy has demonstrated the ability to detect physicochemical attributes in grains with high precision, representing an alternative to conventional destructive analyses (Barnaby *et al.*, 2020). When combined with machine learning and deep learning algorithms, this approach enhances the interpretability of spectral data and enables robust predictive models for physicochemical attributes such as protein, fiber, moisture, lipid, ash, and starch contents (Carneiro *et al.*, 2023). Deep learning-based models have been progressively employed in the prediction of rice composition with good levels of accuracy (Razavi *et al.*, 2024).

Despite the significant advancement of spectroscopic techniques in other agricultural crops such as wheat, corn, and soybean, the consolidated application of these tools in rice still requires further investigation and methodological standardization (Jiang *et al.*, 2023; Teodoro *et al.*, 2021). This gap becomes evident when considering the compositional variability among different types of white, black, red, and parboiled rice, which demand specific analysis and predictive modeling strategies (Chen *et al.*, 2021). Moreover, the development of robust models that integrate these spectral data with physicochemical attributes still faces challenges, especially in contexts requiring high accuracy and interoperability across different datasets (Assadzadeh *et al.*, 2020; Carneiro *et al.*, 2023). In this scenario, the combination of VIS-SWIR spectroscopy and machine learning, comprising shallow and/or deep learning algorithms, emerges as a new alternative. However, further investigation is needed to ensure its full potential in post-harvest and quality control environments. Given the growing demand for faster, more reliable, and non-destructive analytical methods, this study aimed to investigate the potential of VIS and SWIR spectroscopy, combined with shallow and deep learning algorithms, for differentiating various types of rice. The spectral signatures and their relationships with physicochemical attributes and physical defects of the grains were analyzed to develop predictive models supporting classification and quality control.

The purpose of this study was to provide more efficient and non-destructive tools that can be applied from post-harvest stages through industrial processing, thereby adding value to the production chain. The objective was to correlate physicochemical attributes of milled rice using non-destructive techniques coupled with machine learning algorithms. Specifically, the aims were: i) to spectrally characterize white, black, red, and parboiled milled rice using VIS/SWIR spectroscopy; ii) to correlate the physicochemical attributes of the grains with their respective spectral signatures; and iii) to develop and evaluate predictive models based on shallow and deep learning algorithms in a non-destructive approach.

2. MATERIAL AND METHODS

2.1. Sample collection and preparation

The milled rice samples were obtained immediately after milling (without storage) at a processing facility located in the municipality of Cachoeira do Sul, State of Rio Grande do Sul, Brazil, at latitude 30° 0' 45'' S, longitude 52° 55' 11'' W, and an altitude of 73 m. Figure 1 presents the samples of white, parboiled, red, and black rice.

Fig. 1. Samples of milled rice: white rice (a), parboiled rice (b), red rice (c), and black rice (d).

The samples were separated according to Normative Instruction No. 6 of February 16, 2009, based on the technical regulations for rice (Ministry of Agriculture, 2009). Rice samples were composed according to the maximum allowable defect limits for Type 1, based on Table 1 (Brazil. Normative Instruction 2/2012. Ministry of Agriculture, 2012). Each sample consisted of 2 kg of rice, subsequently divided into 100 subsamples of 20 g each. The maximum allowable tolerance for non-parboiled rice grains is 0.30% for all types; any product exceeding this limit is classified as out of grade.

Table 1. Maximum tolerance limits expressed (%)

Rice	Type	Foreign matter and impurities	Moldy and burning	Pitted or stained	Plastered green	Striped	Yellow	Total broken and quirera
White	1	1	0.2	0.5	1.5	1.75	0.3	14
Parboiled	1	1	0.1	0.15	1.5	1.75	0.3	14
Red	1	1	0.2	0.5	1.5	1.75	0.3	14
Black	1	1	0.2	0.5	1.5	1.75	0.3	14

2.2. Near-infrared spectroscopy (NIR)

The determination of the proximate composition-comprising crude protein (CP), moisture content (MC), ether extract (EE), crude fiber (CF), ash (A), and starch (S) was performed using near-infrared spectroscopy (NIR) with high optical precision (Metrohm DS2500 spectrometer, Herisau, Switzerland). All analyses were carried out in triplicate. Homogenized samples were placed in the sampling capsule for scanning. The analysis involved illuminating each sample with radiation at specific wavelengths within the near-infrared region and measuring the difference between the energy emitted by the spectroscope and that reflected by the sample to the detector. This difference was recorded across several bands, generating a spectrum for each sample. Spectral data were collected in reflectance mode, covering a range from 400 to 2 500 nm (Barnes *et al*., 1989). Outputs were compared with a calibration set (Horwitz, 1970). The values obtained from the DS2500 represented the actual physicochemical results for each sample and were used as the reference data in this study. These laboratory-based measurements were later paired with the hyperspectral signatures acquired using the FieldSpec 4 system (350-2 500 nm) to build the regression models. Each physicochemical attribute was treated independently as a separate regression target and was not used to classify the rice types or to define quality thresholds.

2.3. Spectral variables

Spectral analysis was carried out using a spectroradiometer (FieldSpec 4 Jr, Analytical Spectral Devices, Boulder, USA) fitted with a muglight. The equipment measures spectra in the 350-2 500 nm range. The reading interval is 1.4 nm from 350-1 050 and 2 nm from 1 000-2 500 nm. The spectral resolution is 3 nm in the 350-700 nm range and 30 nm in the 1 400-2 100 nm range. One advantage of this method is the preservation of the samples, with minimal influence from environmental factors such as ambient light, thus reducing errors associated with diffuse illumination.

Each spectral reading was performed three times per sample, and the data used for analysis were based on the mean of these three readings. A white reference plate, composed of barium sulfate and reflecting 100% of incident light, was used as a standard. The system stored spectral data from this plate, which were later used to determine the reflectance factor for each sample by multiplying it by the readings obtained for each case. For data analysis, the sensor was connected to a computer equipped with RS³ software for recording measurements, allowing subsequent importation by ViewSpectroPro for data extraction in .txt files and facilitating further statistical analysis.

2.4. Multivariate, prediction, and statistical analyses

For predictive modeling, machine learning analyses were primarily conducted in the Python software using the *scikit-learn, pandas, numpy, matplotlib, seaborn,* and *tensorflow* libraries. Initially, Pearson correlation was applied among the physicochemical variables (moisture, protein, lipids, fiber, starch, and ash) to explore relationships between nutritional attributes. Subsequently, Principal Component Analysis (PCA) was performed separately on the spectral data (350-2 500 nm) and physicochemical variables to visualize clustering patterns among the different rice types.

Machine leaning analysis consisted of the following models: i) shallow algorithms – random forest (RF), gradient boosting (GB), support vector machine (SVM), and the traditional multiple linear regression (LR) model (Table 2); and ii) deep learning models – Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), with architectures consisting of Conv1D and LSTM layers, respectively (Table 3).

Predictions were carried out individually for each physicochemical attribute (output variables), while spectral data were used as input variables. Spectral data were pre-processed by normalization using StandardScaler. Model evaluation was conducted through 10-fold cross-validation, with 10 repetitions applied to traditional models (RF, GB, SVM, LR) to ensure greater statistical robustness. For CNN and RNN models, one-dimensional tensors were used with architectures comprising Conv1D (CNN) and LSTM (RNN) layers. Regularization was performed by applying Dropout (rate of 0.5) and EarlyStopping set with a patience

Table 2. List of machine learning models used in rice quality classification

Acronym	Machine learning model	Reference
RF	Random forest	(Belgiu and Drăguţ, 2016)
SVM	Support vector machine	(Cortes and Vapnik, 1995)
GB	Gradient boosting	(Friedman, 2001)
LR	Multiple linear regression	(Draper and Smith, 1998)
CNN	Convolutional neural networks	(Lecun *et al*., 2002)
RNN	Recurrent neural networks	(Hochreiter and Schmidhuber, 1997)

Table 3. Hyperparameters used for deep learning models

Model	Hiperparameter	Valor
Support vector machine (SVM)	Kernel	rbf
	C	10
	Epsilon	0.2
	Gamma	Automatic (rbf)
Random forest (RF)	n_estimators	100
	max_depth	15
	Bootstrap	True
	Random State	42
Gradient boosting (GB)	n_estimators	100
	learning_rate	0.1
	max_depth	4
	Subsample	1.0 (standard)
	Random State	42
Linear regression (LR)	Solver	auto (standard)
	Normalization	Not applicable
	Intercept	True

Model	Layer type	Number of layers	Neurons	Dropout	Optimizer	Epochs	Batch size
CNN	Conv1D + MaxPooling + Dense	4	64 and 128	0.5	Adam	50	32
RNN	LSTM + Dense	4	64 and 128	0.5	Adam	50	32

of 10 epochs, monitoring the validation loss (val_loss) metric and restoring the best weights at the end of training. Batch size adopted for CNN and RNN was equal to 32.

The correlation coefficient (R), coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE) were used as model performance metrics. Residual analyses (observed vs. predicted values) and graphical visualization via boxplots and histograms were also performed. Furthermore, variable importance analysis was performed using the RF model for each attribute aiming at identifying the most relevant spectral bands. Model performance metrics were represented graphically by boxplots.

The different models were compared based on the evaluated metrics. For this purpose, an analysis of variance (ANOVA) was applied. When the model effect was significant (F-test, $p \leq 0.05$), the means of the performance metrics assessed for each model were grouped by the Scott-Knott test at 5% significance level using Sisvar software v.5.6 (Ferreira, 2019).

The choice of regression algorithms was intentional and grounded in what is already well established in the hyperspectral modeling literature. Shallow learning models such as SVM, RF, and GB are widely used because they deal well with high-dimensional and collinear data, which fits the characteristics of full-spectrum rice measurements. Deep learning models (CNN and RNN) were also included because they can capture local and sequential patterns that sometimes appear in spectral signatures. Together, these approaches offer complementary ways of modeling the physicochemical attributes. The hyperparameters were defined based on recommendations from previous studies and on initial tests performed to reduce overfitting and keep the training behavior stable.

3. RESULTS AND DISCUSSION

3.1. Physicochemical characteristics and spectral signatures by rice type

Table 4 shows the analysis of variance (ANOVA) applied to the main physicochemical characteristics of four rice types: white, parboiled, red, and black. The analyzed physicochemical attributes were moisture, starch, protein, fiber, ash, and lipids. All differences among the rice types were statistically significant, with Pr>Fc = 0.0000 for all variables evaluated.

These findings are consistent with previous studies that highlight black rice as the type with the highest nutritional value among the varieties analyzed (Banerjee *et al.*, 2023; Khatun and Mollah, 2024). Black rice exhibits higher levels of protein, lipids, and phenolic compounds and is frequently classified as a functional food (Zhao *et al.*, 2021; Arashloo and Witter, 2022; Das *et al.*, 2023). This distinctive composition justifies the elevated contents observed in this study, reflecting the presence of bioactive components and minerals concentrated in the hull and pericarp of the grain (Banerjee *et al.*, 2023).

Regarding parboiled rice, the high fiber content identified in this study is consistent with findings from the literature (Akhter *et al.*, 2023). Moreover, the parboiling process promotes the migration of nutrients from the husk to the endosperm (Kalita *et al.*, 2021), increasing the concentration of insoluble fibers. However, this process can also result in the loss of some proteins (Nirmagustina and Handayani, 2023) and lipids (Mudgal and Singh, 2024), which may explain the lower levels of these components found in the parboiled samples evaluated here.

White rice, in turn, exhibited the highest starch content (Yoviono *et al.*, 2022). This type is primarily composed of Starch stored in the endosperm, which constitutes the bulk of the grain. The concentration, structure, and distribution of starch in the endosperm are influenced by genetic, metabolic, and environmental factors (Yang *et al.*, 2020), which in turn affect rice quality, digestibility, and grain development (Cao *et al.*, 2022). This characteristic renders white rice less nutritious compared to other types (Hashimoto *et al.*, 2022; Zhang *et al.*, 2023), yet more palatable and faster to cook, accounting for its popularity in human consumption (Gondal *et al.*, 2021; Zhao *et al.*, 2021).

Red rice exhibited an intermediate composition for seve-ral characteristics, but stood out negatively for its fiber content, which was the lowest among the types evaluated. These results diverge from those found by Abeysiriwardena and Gunasekara (2020) and Gogoi *et al.* (2025), who highlight its high dietary fiber content. Although red rice is recognized for its anthocyanin-rich coloration, its macronutrient values can vary widely depending on genetic variety and agricultural management practices (Gogoi *et al.*, 2025).

These differences highlight the distinct functional and nutritional potential among rice types and justify the growing interest in using spectroscopy and predictive modeling as tools for rapid and non-destructive differentiation of nutritional quality attributes (Liu *et al.*, 2020; Johnson *et al.*, 2021; Xu *et al.*, 2025). The coefficients of variation (CV%) ranged between 7.26 and 13.39%, all within acceptable limits, indicating good data homogeneity.

Figure 2 presents boxplots visually depicting the distribution of the numerical values for the physicochemical characteristics across the four rice types. Marked differen ces between the groups are evident, graphically confirming the ANOVA results. Asymmetric distributions are observed for some characteristics, such as fiber in parboiled rice, indicating possible variability in the material. Outlier values generally appear discretely and do not compromise the analysis or visualization, allowing it to be observed that black rice distinctly stands out in variables such as protein, lipids, and ash.

Figure 2, in addition to Table 4, highlights the clear chemical differentiation among the rice types, again emphasizing black rice for its higher values in protein and lipids, and parboiled rice for its high fiber content. These results reinforce the importance of classifying rice according to its nutritional properties.

Table 4. Analysis of variance of the physicochemical characteristics of different types of rice

Processes	Moisture	Starch	Protein	Lipids	Fiber	Ashes
	(%)					
White rice	11.12 b	73.39 a	8.04 b	1.54 c	2.10 b	1.12 c
Parboiled rice	11.19 b	70.11 b	4.19 d	1.44 d	2.80 a	1.23 b
Red rice	13.16 a	66.40 c	6.88 c	1.94 b	0.69 d	1,02 d
Black rice	13.35 a	57.06 d	9.08 a	2.04 a	1.18 c	1,43 a
Pr>Fc	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
CV (%)	8.37	7.26	10.38	13.39	10.63	8.67
General average	12.21	66.74	7.05	1.74	1.69	1.2

The asterisk (*) was explained as indicating statistical significance at $p \leq 0.05$ by the F test (ANOVA).

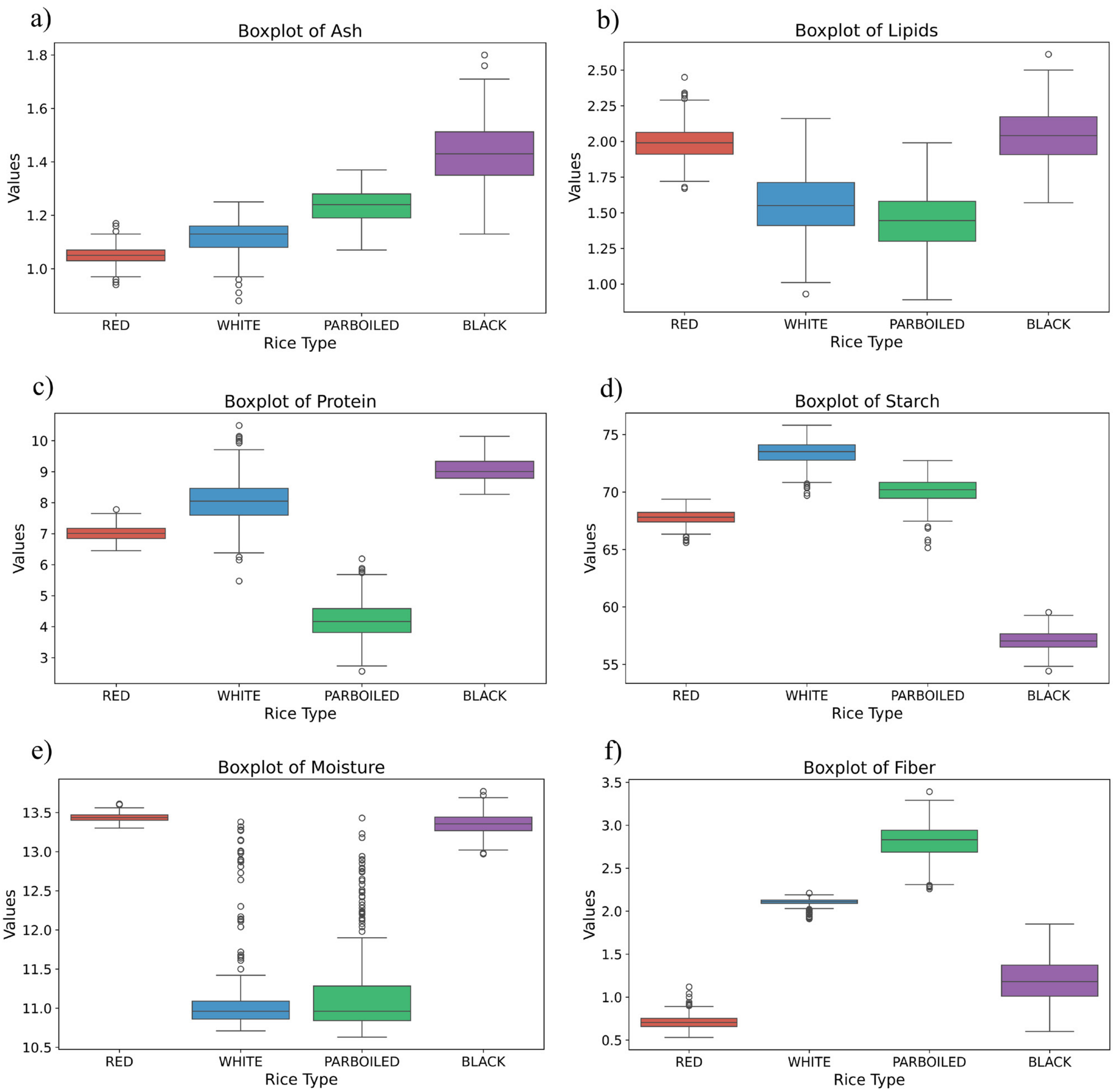

Fig. 2. Boxplots of physicochemical characteristics for red, white, black, and parboiled rice types.

The PCA for the physicochemical characteristics (Fig. 3) considers variables such as moisture, protein, lipids, fiber, starch, and ash, which are laboratory-derived attributes of the grains. The analysis of the plot reveals partial clustering among the rice types, where black rice tends to be distinctly separated from the others, possibly due to its higher lipid and protein contents, as previously shown in Fig. 2. Red and white rice exhibit greater overlap, indicating that, based solely on physicochemical variables, there is greater similarity between these two types. This supports the findings shown by Pereira *et al.* (2023), who reported that traditional physicochemical characteristics, though relevant, are not always sufficient to clearly distinguish rice varieties with intermediate morphology and composition. While good dispersion among the samples is noted-suggesting a reasonable degree of differentiation – it is not sufficiently strong, especially when compared to the spectral data, to guarantee clear separation of all rice types. Although useful, physicochemical PCA demonstrates limitations in distinguishing rice types, particularly those more similar in laboratory attributes (Gao *et al.*, 2024; Hazrul *et al.*, 2025).

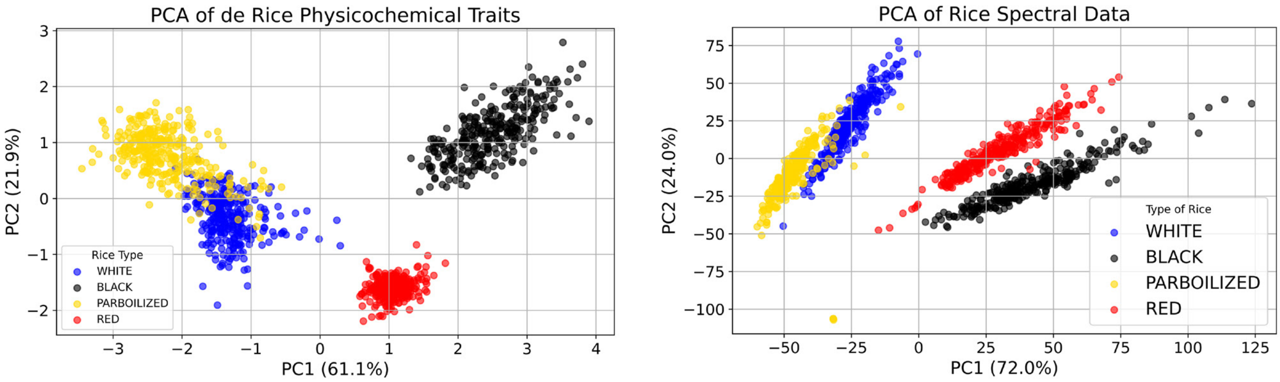

Fig. 3. Principal component analysis (pca) of spectral data collected from white, red, black, and parboiled rice types.

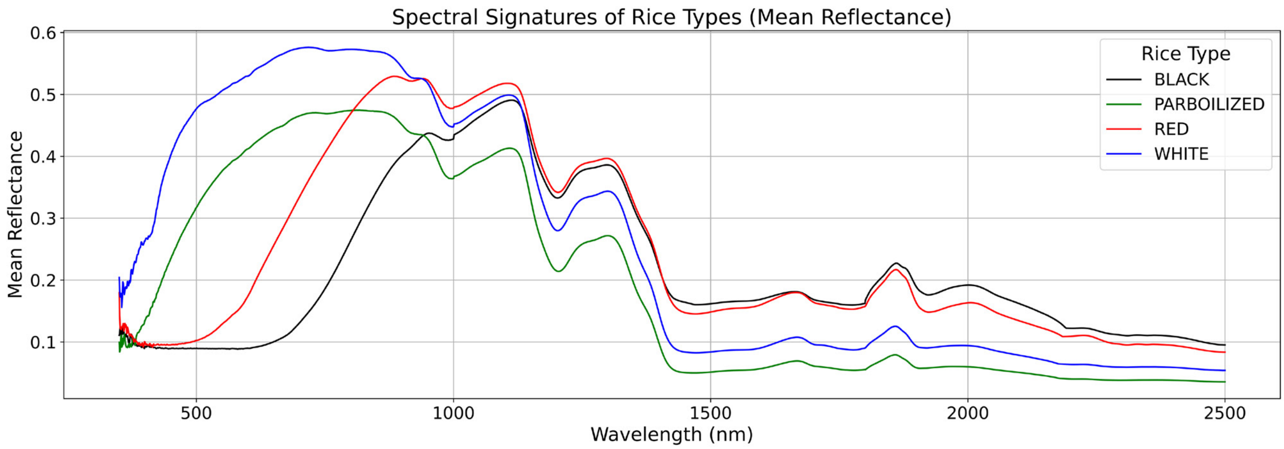

Fig. 4. Mean spectral signatures of black, red, white, and parboiled rice types.

In contrast, the PCA for Spectral Characteristics, which considers spectral bands from 350 to 2500 nm, shows a markedly greater discrimination between groups. In this PCA, unlike the physicochemical one, there is a clearer separation among the rice types. Each group – black, red, white, and parboiled rice – tends to cluster in distinct regions of the plot. This indicates that spectral signatures capture specific structural and compositional features of each type, which are often imperceptible in laboratory analyses. Thus, spectral PCA provides better discrimination among rice types, demonstrating that spectral data are more sensitive to grain composition than traditional physicochemical data. This sensitivity of spectral data to grain composition was also reported by Kang *et al.* (2024) when discriminating among other rice varieties.

This analysis may be crucial in the post-harvest context, where decisions impact quality preservation, market value, and operational efficiency (Mahmood *et al.*, 2024; Müller *et al.*, 2022). Therefore, PCA not only contributes to subsequent predictive models but also reinforces the role of non-destructive technologies as screening and real-time quality control tools in post-harvest processes.

3.2. Spectral signatures of rice types

Figure 4 shows the mean spectral reflectance curves for black, red, white, and parboiled rice, covering the spectral range from 350 to 2500 nm. The curves exhibit well-defined patterns among the different rice types, particularly in the VIS (350-700 nm) and SWIR (700-2500 nm) regions, confirming the capability of spectroscopy to differentiate the groups.

Regarding black rice, as shown in Fig. 4, lower reflectance values are observed throughout the entire spectrum, especially in the visible region. The visible color of black rice results from light reflected off the bran layer, which is rich in anthocyanins; thus, the higher the anthocyanin concentration, the darker and more intense the appearance. Darker black rice grains contain higher concentrations of these beneficial compounds compared to lighter or white grains (Theiventhiran *et al.*, 2020; Brunet-Loredo *et al.*, 2023; Borah *et al.*, 2025). White rice, in turn, exhibits higher overall reflectance, particularly in the VIS and transition regions with the SWIR, which is expected due to its lighter coloration (García-Salcedo *et al.*, 2023; Aekram *et al.*, 2025). Red and parboiled rice occupy intermediate positions, presenting relatively similar signatures, though subtle differences are noticeable in specific regions, such as between 1 100-1 400 nm and 1 900-2 100 nm. This visualization reinforces that spectral signatures are unique to each rice type, reflecting their distinct spectral and structural characteristics.

These patterns support the use of machine learning models and spectral analysis for the prediction of nutritional attributes. Such findings are particularly relevant in the post-harvest context, as they enable the rapid and non-destructive screening of rice lots, optimizing classification processes and the commercial targeting of grains. The figure also provides visual evidence complementary to Fig. 1, emphasizing that rice types exhibit distinct signatures, thus enabling their identification.

3.3. Evaluation of shallow and deep learning models

As presented in Fig. 5, the shallow learning models SVM, RF, and GB consistently formed the best-performing group in almost all physicochemical attributes. These models reached the highest R and R^2 values and the lowest MAE and RMSE, remaining statistically similar in the ANOVA and Scott-Knott tests shown in Tables 5-10. For example, in the prediction of protein content in Fig. 5e, SVM reached R = 0.952 and R^2 = 0.904, followed closely by RF and GB with equally high accuracy.

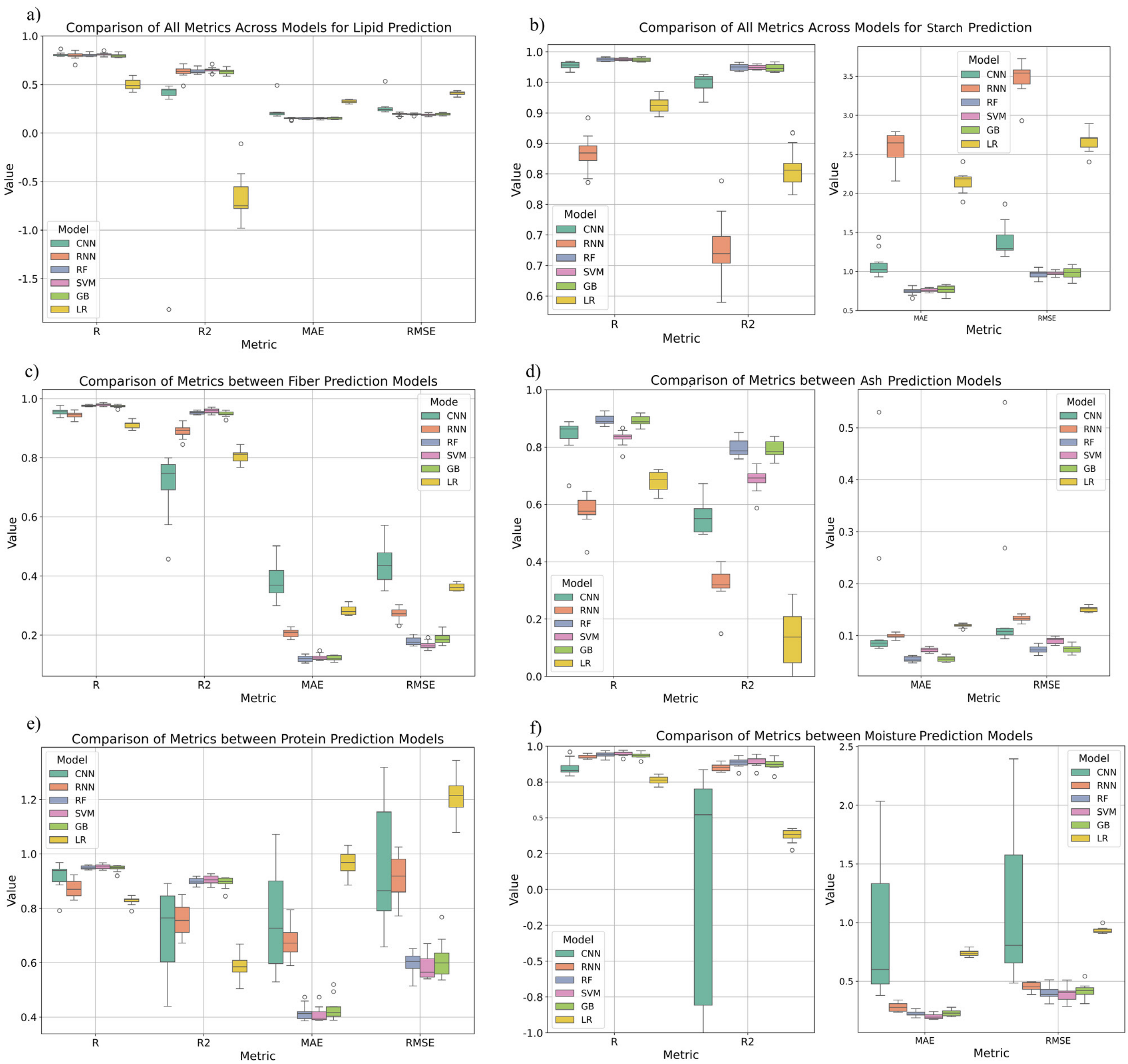

Fig. 5. Boxplots of R, R^2, MAE, and RMSE for all models for each rice attribute: a) lipids, b) starch, c) moisture, d) ash, e) protein, f) fiber.

Table 5. Analysis of variance of machine learning models for the different types of rice in protein prediction

Model	R	R^2	MAE	RMSE
LR	0.826970 a	0.584926 a	0.964798 a	1.212073 a
RNN	0.875092 b	0.760400 b	0.679102 c	0.911554 b
CNN	0.917549 c	0.719705 b	0.756813 b	0.956924 b
GB	0.946986 d	0.894261 c	0.431075 d	0.611450 c
RF	0.949667 d	0.899362 c	0.416296 d	0.597104 c
SVM	0.952401 d	0.903906 c	0.409373 d	0.581563 c
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	2.82	8.58	13.83	14.06
Overall mean	0.9114440	0.7937598	0.6095761	0.8117780

*Explanation as in Table 4.

Table 6. Analysis of variance of machine learning models for the different types of rice in starch prediction

Model	R	R^2	MAE	RMSE
RNN	0.833538 a	0.677346 a	2.585056 d	3.468162 d
LR	0.913455 b	0.808589 a	2.151020 c	2.671538 c
CNN	0.977529 c	0.947339 c	1.088146 b	1.394123 b
GB	0.986908 c	0.973679 d	0.766993 a	0.984199 a
RF	0.987449 c	0.974739 d	0.748048 a	0.966918 a
SVM	0.987520 c	0.974626 d	0.761022 a	0.972568 a
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	1.35	2.75	8.97	7.96
Overall mean	0.9477333	0.8927196	1.3500474	1.7429178

*Explanation as in Table 4.

Table 7. Analysis of variance of machine learning models for the different types of rice in moisture prediction

Model	R	R^2	MAE	RMSE
LR	0.761479 a	0.376972 b	0.740377 b	0.933110 b
CNN	0.850031 b	-0.232026 a	0.912414 b	1.122891 b
RNN	0.927148 c	0.852417 c	0.280969 a	0.452444 a
GB	0.935030 c	0.872243 c	0.229157 a	0.418233 a
RF	0.941759 c	0.884839 c	0.223402 a	0.397731 a
SVM	0.945649 c	0.889280 c	0.199913 a	0.388168 a
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	2.59	97.80	60.45	45.55
Overall mean	0.8935161	0.6072874	0.4310389	0.6187627

*Explanation as in Table 4.

Table 8. Analysis of variance of machine learning models for the different types of rice in fiber prediction

Model	R	R^2	MAE	RMSE
LR	0.912111 a	0.806076 a	0.284225 b	0.362201 b
RNN	0.944472 b	0.890028 b	0.206893 c	0.271541 c
CNN	0.954765 c	0.704363 c	0.379771 a	0.442052 a
GB	0.974320 d	0.948111 d	0.121865 d	0.187390 d
RF	0.976409 d	0.952487 d	0.120353 d	0.179621 d
SVM	0.979788 d	0.958903 d	0.125506 d	0.165917 d
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	0.87	5.24	13.35	12.34
Overall mean	0.9569776	0.8766614	0.2064355	0.2681203

*Explanation as in Table 4.

Table 9. Analysis of variance of machine learning models for the different types of rice in ash prediction

Model	R	R^2	MAE	RMSE
RNN	0.479685 a	0.224316 a	0.108010 b	0.141589 b
LR	0.680372 b	0.111755 b	0.118993 b	0.151002 b
SVM	0.831289 c	0.685680 d	0.072255 a	0.090049 a
CNN	0.845071 c	0.511615 c	0.115809 b	0.137690 b
GB	0.891771 d	0.791237 e	0.054818 a	0.073521 a
RF	0.894952 d	0.797437 e	0.054389 a	0.072392 a
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	4.44	12.69	39.03	30.43
Overall mean	0.7705233	0.5203401	0.0873789	0.1110405

*Explanation as in Table 4.

Table 10. Analysis of variance of machine learning models for the different types of rice in the prediction of lipids

Model	R	R^2	MAE	RMSE
LR	0.498860 a	-0.651551 a	0.325620 c	0.410776 c
RNN	0.794359 b	0.628090 c	0.150674 a	0.195300 a
GB	0.799374 b	0.632075 c	0.153244 a	0.194911 a
RF	0.801374 b	0.636410 c	0.150537 a	0.193632 a
CNN	0.809633 b	0.430318 b	0.226324 b	0.272005 b
SVM	0.810526 b	0.650851	0.150325 a	0.189891 a
Prob>F	0.0000*	0.0000*	0.0000*	0.0000*
CV (%) =	4.01	27.83	20.49	16.21
Overall mean	0.7523545	0.3876988	0.1927873	0.2427526

*Explanation as in Table 4.

The overall pattern was stable for starch, moisture, and fiber in Fig. 5b, 5c, and 5f, reinforcing the robustness of shallow learning approaches for this hyperspectral dataset. Two exceptions were observed. In the lipid prediction shown in Fig. 5a, RNN presented performance like SVM, RF, and GB, joining the same statistical group. For ash in Fig. 5d, the separation of statistical groups differed slightly from the other attributes, although SVM, RF, and GB still presented the best absolute accuracy.

In general, the deep learning models CNN and RNN showed greater variability and lower predictive stability compared to shallow models, which were expected to be given the moderate size of the dataset. Overall, the shallow algorithms presented the most consistent and reliable performance across all rice attributes.

This behavior reinforces recent findings in the scientific literature that highlight the robustness of models like SVM and RF for regression tasks with hyperspectral data (Feyisa *et al.*, 2020; Liu *et al.*, 2022; Tunca *et al.*, 2023), especially in contexts involving moderate-sized datasets without significant temporal variation (Li *et al.*, 2023; Nagy and Neff 2024). This is partly due to the ability of these algorithms to capture nonlinear relationships in the data without depending on complex adjustments, such as deep hidden layers or extended training sequences. In addition, the GB model proved to be a relevant tool for predicting physicochemical properties in maize (Zheng *et al.*, 2024; Zhao *et al.*, 2025) and soybean (Huber *et al.*, 2022; Li *et al.*, 2023) using complex spectroscopic data – findings that were also replicated in this study.

Deep learning models, although promising, demonstrated greater variability across folds and were more susceptible to overfitting or required finer hyperparameter tuning. For instance, the CNN exhibited inferior performance across almost all analyzed variables, likely due to its limited ability to capture the sequential dependencies present in spectral data. The RNN performed slightly better, as it is more suited to sequence modeling, but still lagged behind traditional models – a result that may be attributed to the relatively small dataset, a common limitation in experimental studies with physical samples (Liu *et al.*, 2024).

These findings reinforce that, for the context evaluated here – predicting physicochemical properties from hyperspectral data – classical machine learning models remain among the most efficient approaches (Carneiro *et al.*, 2023), both for their precision and statistical consistency, demonstrating that they are highly effective tools in various agricultural scenarios.

4. CONCLUSIONS

This study provides evidence of the effectiveness of integrating hyperspectral spectroscopy (350-2 500 nm) and machine learning algorithms for predicting the physicochemical attributes of different rice types, including protein, starch, moisture, fiber, ash, and lipids. Among the evaluated models, SVM exhibited superior performance, demonstrating statistical robustness and strong generalization ability across the variables. The detailed analysis of spectral bands enabled the identification of representative regions for prediction, reinforcing the potential for developing optimized multispectral sensors. Overall, the findings consolidate the application of non-destructive models in the post-harvest context, offering advancements for the optimization of industrial operations as well as improvements in accuracy for classification and quality control processes.

Declaration of Competing Interest: The authors declare that there is no conflict of interest in the research.

5. ACKNOWLEDGEMENT

The authors would like to thank UFSM (Federal University of Santa Maria) Laboratory of Postharvest (LAPOS) Research Group at Postharvest Innovation: Technology, Quality and Sustainability, UPF (University of Passo Fundo) Cereal Laboratory, and Federal University of Mato Grosso do Sul (UFMS) for their contributions in the research project.

6. REFERENCES

Abeysiriwardena, D.S.D.Z., Gunasekara, D.C.S., 2020. Development of a red rice variety with excellent health properties and attractive grain qualities. Indian J. Genetics Plant Breeding 80(01), 115-117. https://isgpb.org/journal/index.php/IJGPB/article/view/116

Aekram, S., Klinsuk, J., Budtri, N., Sunongbua, P., Somphong, C., Lertsiriyothin, W., 2025. Differentiation of white rice Khao Dawk Mali 105 (KDML105) from other Thai white rice using FTIR microscope-attenuated total reflectance-focal plane array detector. Food Res. 9(1), 282-287. https://doi.org/10.26656/fr.2017.9(1).571

Akhter, K.T., Shozib, H.B., Islam, M.H., Sarwar, S., Islam, M.M., Akanda, M.R., *et al.*, 2023. Variations in the major nutrient composition of dominant high-yield varieties (hyvs) in parboiled and polished rice of Bangladesh. Foods 12(21), 3997. https://doi.org/10.3390/foods12213997

Arashloo, S.R., Kittler, J., 2022. Multi-target regression via nonlinear output structure learning. Neurocomputing 492, 572-580. https://doi.org/10.1016/j.neucom.2021.12.048

Assadzadeh, S., Walker, C.K., McDonald, L.S., Maharjan, P., Panozzo, J.F., 2020. Multi-task deep learning of near infrared spectra for improved grain quality trait predictions. J. Near Infrared Spectroscopy 28(5-6), 275-286. https://doi.org/10.1177/0967033520939318

Banerjee, R., Das Gupta, B., Kar, A., Bhardwaj, P.K., Sharma, N., Haldar, P.K., *et al.*, 2023. Quality evaluation of different black rice varieties of northeastern region of India. Phytochemical Analysis 34(5), 507-517. https://doi.org/10.1002/pca.3230

Barnaby, J.Y., Huggins, T.D., Lee, H., McClung, A.M., Pinson, S.R., Oh, M., *et al.*, 2020. Vis/NIR hyperspectral imaging distinguishes sub-population, production environment, and physicochemical grain properties in rice. Scientific reports 10(1), 9284. https://doi.org/10.1038/s41598-020-65999-7

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Applied Spectroscopy 43(5), 772-777. https://doi.org/10.1366/0003702894202201

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogrammetry Remote Sensing, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Borah, J.L., Tampha, S., Jose, R.C., Chaudhary, S.K., Chetia, S.K., Tamuly, A., *et al*., 2025. A validated method for identification and quantification of anthocyanins in different black rice (*Oryza sativa* L.) varieties using high-performance thin-layer chromatography (HPTLC). Phytochemical Analysis 36(4) 1268-1277. https://doi.org/10.1002/pca.3510

Brazil. Normative instruction 2/2012. Ministry of Agriculture, p. and supply. 2012. Brazil. 2/2012. Brazil: [s.n.].

Brazil, Ministry of Agriculture, Livestock and Food Supply. Normative Instruction No. 6, of February 16, 2011. Official rice classification standards. Official Gazette of the Union: section 1, Brasília, DF, February 17, 2011.

Brotman, Y., Llorente-Wiegand, C., Oyong, G., Badoni, S., Misra, G., Anacleto, R., *et al*., 2021. The genetics underlying metabolic signatures in a brown rice diversity panel and their vital role in human nutrition. Plant J. 106(2) 507-525. https://doi.org/10.1111/tpj.15182

Brunet-Loredo, A., López-Belchí, M.D., Cordero-Lara, K., Noriega, F., Cabeza, R.A., Fischer, S., *et al*., 2023. Assessing grain quality changes in white and black rice under water deficit. Plants 12(24), 4091. https://doi.org/10.3390/plants12244091

Cao, R., Zhao, S., Jiao, G., Duan, Y., Ma, L., Dong, N., *et al*., 2022. OPAQUE3, encoding a transmembrane bZIP transcription factor, regulates endosperm storage protein and starch biosynthesis in rice. Plant Communications 3(6), 100463. https://doi.org/10.1016/j.xplc.2022.100463

Carneiro, L.O., Coradi, P.C., Rodrigues, D.M., Lima, R.E., Teodoro, L.P.R., Santos de Moraes, R., *et al*., 2023. Characterizing and predicting the quality of milled rice grains using machine learning models. Agri. Eng. 5(3), 1196-1215. https://doi.org/10.3390/agriengineering5030076

Chen, H., Chen, D., He, L., Wang, T., Lu, H., Yang, F., *et al*., 2021. Correlation of taste values with chemical compositions and rapid visco analyser profiles of 36 indica rice (*Oryza sativa* L.) varieties. Food Chemistry 349, 129176. https://doi.org/10.1016/j.foodchem.2021.129176

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning, 20, 273-297. https://doi.org/10.1007/BF00994018

Das, M., Dash, U., Mahanand, S.S., Nayak, P.K., Kesavan, R.K., 2023. Black rice: A comprehensive review on its bioactive compounds, potential health benefits and food applications. Food Chem. Advances 3, 100462. https://doi.org/10.1016/j.focha.2023.100462

Draper, N.R., Smith, H., 1998. Applied Regression Analysis (Vol. 326). Wiley and Sons.

Ferreira, D.F., 2019. SISVAR: A computer analysis system to fixed effects split plot type designs. Brazilian J. Biometrics 37(4), 529-535. https://doi.org/10.28951/rbb.v37i4.450

Feyisa, G.L., Palao, L.K., Nelson, A., Gumma, M.K., Paliwal, A., Win, K.T., *et al*., 2020. Characterizing and mapping cropping patterns in a complex agro-ecosystem: An iterative participatory mapping procedure using machine learning algorithms and MODIS vegetation indices. Computers Electronics in Agric. 175, 105595. https://doi.org/10.1016/j.compag.2020.105595

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29(5), 1189-123. https://www.jstor.org/stable/2699986

Gao, Y., Zhang, L., Chen, W., Zhou, W., Deng, G., Dai, G., *et al*., 2024. Cooked Rice textural properties and starch physicochemical properties from new hybrid Rice and their parents. Foods 13(7), 1035. https://doi.org/10.3390/foods13071035

García-Salcedo, Á.J., Fonthal, G., Franco, J.F., Tirado-Mejía, L., 2023. Analysis of compositional differences between commercial rice grains by the study of the photoluminescence response. J. Cereal Sci. 111, 103681. https://doi.org/10.1016/j.jcs.2023.103681

Gogoi, P., Sharma, P., Goudar, G., Mahajan, A., Dutta, H., Sasikumar, R., *et al*., 2025. Nutritional profile and mineral bioaccessibility of pigmented rice landraces. J. Food Measurement Characterization, 19(2), 1513-1530. https://doi.org/10.1007/s11694-024-03094-5

Gondal, T.A., Keast, R.S., Shellie, R.A., Jadhav, S.R., Gamlath, S., Mohebbi, M., *et al*., 2021. Consumer acceptance of brown and white rice varieties. Foods 10(8), 1950. https://doi.org/10.3390/foods10081950

Graves, A., Graves, A., 2012. Long short-term memory. supervised sequence labelling with recurrent neural networks. Springer Nature 37-45. https://doi.org/10.1007/978-3-642-24797-2_4

Grinsztajn, L., Oyallon, E., Varoquaux, G., 2022. Why do tree-based models still outperform deep learning on typical tabular data? Advances in Neural Information Processing Systems, 35, 507-520. https://proceedings.neurips.cc/paper_files/paper/2022/hash/0378c7692da36807bdec87ab043cdadc-Abstract-Datasets_and_Benchmarks.html

Hashimoto, M., Hossain, S., Matsuzaki, K., Shido, O., Yoshino, K., 2022. The journey from white rice to ultra-high hydrostatic pressurized brown rice: an excellent endeavor for ideal nutrition from staple food. Critical Reviews in Food Science and Nutrition, 62(6), 1502-1520. https://doi.org/10.1080/10408398.2020.1844138

Hazrul, A.F.H., Ibrahim, R.R., Duralim, M., 2025. Discrimination of rice varieties using laser-induced breakdown spectroscopy assisted with principal component analysis (PCA). In J. Physics: Conf. Series 2974(1), 012009. IOP Publishing. https://doi.org/10.1088/1742-6596/2974/1/012009

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horwitz, W.C.P.R.H., 1970. Official methods of analysis of the Association of Official Analytical Chemists. Washington, DC, USA: Association of Official Analytical Chemists.

Huber, J., Inoua, S., Kerschbamer, R., König-Kersting, C., Palan, S., Smith, V.L., 2022. Nobel and novice: Author prominence affects peer review. Proc. National Academy of Sciences 119(41), e2205779119. https://doi.org/10.1073/pnas.2205779119

Jiang, Q., Zhang, M., Mujumdar, A.S., Wang, D., 2023. Non-destructive quality determination of frozen food using NIR spectroscopy-based machine learning and predictive modelling. J. Food Eng. 343, 111374. https://doi.org/10.1016/j.jfoodeng.2022.111374

Johnson, J.M., Sila, A., Senthilkumar, K., Shepherd, K.D., Saito, K., 2021. Application of infrared spectroscopy for estimation of concentrations of macro-and micronutrients in rice in sub-Saharan Africa. Field Crops Res. 270, 108222. https://doi.org/10.1016/j.fcr.2021.108222

Julcarima, C.B., Rojas Llanos, E.M., 2018. Effect of parboiled on sensorial properties, vitamin and mineral content in rice varieties (*Oryza sativa*) produced in the country, thesis. 1-142, Cerro De Pasco, Peru.

Kabir, M.S., Kanan, M.K.A., Ahmed, J., Hossain, M.A., Akter, M.S., Ahmed, K.S., *et al*., 2024. Free and Insoluble-Bound Phenolics in Parboiled and Non-Parboiled Rice after Hydrothermal Treatments. Maruf, Free and Insoluble-Bound Phenolics in Parboiled and Non-Parboiled Rice after Hydrothermal Treatments Plants 8(4). http://dx.doi.org/10.2139/ssrn.5022123

Kalita, T., Gohain, U.P., Hazarika, J., 2021. Effect of different processing methods on the nutritional value of rice. Current Research in Nutrition and Food Sci. J. 9(2), 683-691. https://dx.doi.org/10.12944/CRNFSJ.9.2.31

Kang, Z., Fan, R., Zhan, C., Wu, Y., Lin, Y., Li, K., *et al*., 2024. The Rapid non-destructive differentiation of different varieties of rice by fluorescence hyperspectral technology combined with machine learning. Molecules 29(3), 682. https://doi.org/10.3390/molecules29030682

Khatun, S., Mollah, M.M.I., 2024. Analysis of black rice and some other cereal grains for protein, sugar, polyphenols, antioxidant and anti-inflammatory properties. J. Agric. Food Res., 16, 101121. https://doi.org/10.1016/j.jafr.2024.101121

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 2002. Gradient-based learning applied to document recognition. Proc. IEEE, 86(11), 2278-2324. https://doi.org/10.1109/5.726791

Lin, H., Jiang, H., Lin, J., Chen, Q., Ali, S., Teng, S., *et al*., 2021. Rice freshness identification based on visible near-infrared spectroscopy and colorimetric sensor array. Food Analytical Methods 14, 1305-1314. https://doi.org/10.1007/s12161-021-01963-z

Liu, X., Wang, W., Li, Z., Xu, L., Lan, D., Wang, Y., 2024. Lipidomics analysis unveils the dynamic alterations of lipid degradation in rice bran during storage. Food Res. Int. 184, 114243. https://doi.org/10.1016/j.foodres.2024.114243

Liu, Y., Li, Y., Peng, Y., Yang, Y., Wang, Q., 2020. Detection of fraud in high-quality rice by near-infrared spectroscopy. J. Food Sci. 85(9), 2773-2782. https://doi.org/10.1111/1750-3841.15314

Liu, Y., Feng, H., Yue, J., Fan, Y., Jin, X., Zhao, Y., *et al*., 2022. Estimation of potato above-ground biomass using UAV-based hyperspectral images and machine-learning regression. Remote Sensing 14(21), 5449. https://doi.org/10.3390/rs14215449

Mahmood, N., Liu, Y., Zheng, X., Munir, Z., Pandiselvam, R., Zhang, Y., *et al*., 2024. Influences of emerging drying technologies on rice quality. Food Res. Int. 184, 114264. https://doi.org/10.1016/j.foodres.2024.114264

Ministry of Agriculture, P.E.A. Approves the technical regulation for rice. Official Gazette of the Union, Brasília. Normative Instruction no. 6. Brazil: [s.n.].

Mudgal, S., Singh, N., 2024. Effect of parboiling treatment times on the physicochemical, cooking, textural, and pasting properties and amino acid, phenolic, and sugar profiles of germinated paddy rice from different rice varieties. J. Food Sci. 89(6), 3208-3229. https://doi.org/10.1111/1750-3841.17048

Müller, A., Nunes, M.T., Maldaner, V., Coradi, P.C., de Moraes, R.S., Martens, S., *et al*., 2022. Rice drying, storage and processing: effects of post-harvest operations on grain quality. Rice Sci. 29(1), 16-30. https://doi.org/10.1016/j.rsci.2021.12.002

Nagy, P., Neff, G., 2024. Conjuring algorithms: Understanding the tech industry as stage magicians. New Media Soc. 26(9), 4938-4954. https://doi.org/10.1177/14614448241251789

Nirmagustina, D.E., Handayani, S., 2023. Nutritional Content of Brown Rice and White Rice from Organic Rice of Mentik Susu Varieties with Parboiled Method. Int. J. Life Sci. Agric. Res. 2(11), 1. https://doi.org/10.55677/ijlsar/V02I11Y2023-01

Pereira, C.L., Sousa, I., Lourenço, V.M., Sampaio, P., Gárzon, R., Rosell, C.M., *et al*., 2023. Relationship between physicochemical and cooking quality parameters with estimated glycaemic index of rice varieties. Foods 13(1), 135. https://doi.org/10.3390/foods13010135

Razavi, M., Mavaddati, S., Koohi, H., 2024. ResNet deep models and transfer learning technique for classification and quality detection of rice cultivars. Expert Systems With Aplications 247, 123276. https://doi.org/10.1016/j.eswa.2024.123276

Teodoro, P.E., Teodoro, L.P.R., Baio, F.H.R., da Silva Junior, C.A., dos Santos, R.G., Ramos, A.P.M., *et al*., 2021. Predicting days to maturity, plant height, and grain yield in soybean: A machine and deep learning approach using multispectral data. Remote Sensing 13(22), 4632. https://doi.org/10.3390/rs13224632

Theiventhiran, T., Vasantha, S., Amutha, J., Ramalingam, G., Hemalatha, Umamaheswari, T., 2020. Screening out the phytochemicals and colour values in white and black rice lines and its interrelationship. Int. J. Biochemistry Res. Revi. 29(3), 34-41. https://doi.org/10.9734/ijbcrr/2020/v29i330178

Tunca, E., Köksal, E.S., Öztürk, E., Akay, H., Çetin Taner, S., 2023. Accurate estimation of sorghum crop water content under different water stress levels using machine learning and hyperspectral data. Environ. Monitoring Assessment 195(7), 877. https://doi.org/10.1007/s10661-023-11536-8

Xu, Y., Dong, Y., Liu, J., Wang, C., Li, Z., 2025. Combination of near infrared spectroscopy with characteristic interval selection for rapid detection of rice protein content. J. Food Composition Analysis 137, 106995. https://doi.org/10.1016/j.jfca.2024.106995

Yang, Y., Lin, G., Yu, X., Wu, Y., Xiong, F., 2020. Rice starch accumulation at different endosperm regions and physical properties under nitrogen treatment at panicle initiation stage. Int. J. Biological Macromolecules 160, 328-339. https://doi.org/10.1016/j.ijbiomac.2020.05.210

Yoviono, F., Sandra, Y., Arifandi, F., 2022. Comparison of starch content in black rice compared to white rice using the iodide test (in Indonesian). Cerdika: Jurnal Ilmiah Indonesia, 2(11), 976-981. http://cerdika.publikasiindonesia.id/index.php/cerdika/index

Zhang, Y., Li, F., Huang, K., Li, S., Cao, H., Xie, J., *et al*., 2023. Structural changes of starch under different milling degrees affect the cooking and textural properties of rice. Food Chemistry 10(17), 100627. https://doi.org/10.1016/j.fochx.2023.100627

Zhao, M., Huang, C., Mao, Q., Zhou, Z., Xiang, Y., Zhang, Q., *et al*., 2021. How anthocyanin biosynthesis affects nutritional value and anti-inflammatory effect of black rice. J. Cereal Sci. 101, 103295. https://doi.org/10.1016/j.jcs.2021.103295

Zheng, W., Zheng, K., Gao, L., Zhangzhong, L., Lan, R., Xu, L., Yu, J., 2024. GRU-transformer: a novel hybrid model for predicting soil moisture content in root zones. Agronomy 14(3), 432. https://doi.org/10.3390/agronomy14030432

Zhao, X., Zhang, Q., Fang, X., Zhang, J., Feng, W., Zhang, L., Wu, X., 2025. Finite element analysis of corn kernel damage: a viscoelastic model approach for impact-induced damage prediction. J. Sci. Food Agric. 105(7), 3993-4004. https://doi.org/10.1002/jsfa.14240