J.Natn.Sci.Foundation Sri Lanka 2025 53 (4): 333 - 343
DOI: http://dx.doi.org/10.4038/jnsfsr.v53i4.12124

RESEARCH ARTICLE

Postharvest Technology

Use of near infrared spectroscopy for rapid and non-destructive identification of morphologically similar seeds of different paddy varieties [*Oryza sativa* L.]

KW Yasintha[1*] and BM Jinendra[2]
[1] *Department of Applied and Natural Sciences, University College of Matara, University of Vocational Techonlogy Sri Lanka.*
[2] *Department of Agricultural Engineering, Faculty of Agriculture, University of Ruhuna, Mapalana, Kamburupitiya, Sri Lanka.*

Submitted: 21 May 2024; Revised: 28 September 2025; Accepted: 29 September 2025

Abstract: Rice production systems and breeding programmes often require handling large numbers of paddy samples, where maintaining sample identity is critical. Conventional identification methods, such as morphological and molecular analyses, are reliable but laborious and time-consuming. To address this, the present study evaluated the use of near-infrared (NIR) spectroscopy (588–1091 nm) combined with soft independent modeling of class analogy (SIMCA) for rapid discrimination of paddy seed varieties. A total of 750 seed samples representing five varieties (Bg300, Bg352, Bg357, Bg359, and Bw363) were scanned, and SIMCA prediction models were developed using Pirouette 3.11 software. Model parameters were optimized to enhance classification performance. The influence of moisture content (MC) at four levels (10%, 12%, 15%, and 23%) during seed handling was also investigated. The SIMCA models achieved variety-specific classification accuracies of 89% (Bg300), 83% (Bg352), 88% (Bg357), 87% (Bg359), and 84% (Bw363). When averaged across all varieties, correct identification rates were 82% at 10% MC, 83% at 12% MC, 88% at 15% MC, and 92% at 23% MC. The results confirm an increasing trend in predictive accuracy with elevated MC, attributable to stronger variety-specific water-related absorbance features in the NIR region. Overall, the study demonstrates that NIR spectroscopy combined with SIMCA provides a rapid, non-destructive, and feasible approach for varietal identification of paddy seeds, particularly effective under equalized moisture levels. This method holds strong potential for field-level application, offering faster and more reliable variety authentication compared to conventional practices.

Keywords: NIR spectroscopy; morphologically similar paddy varieties; *Oryza sativa* L.; principal component analysis; soft independent modeling of class analogy.

INTRODUCTION

In everyday rice breeding trials, within rice breeding stations and in processing systems such as large scale warehouses worldwide, there are hundreds of paddy samples from different varieties to be handled. Exact identification of the correct variety to be used in breeding is essential since an error may cause a significant impact on results and ultimately ends up with wrong decisions, wasting time and money. The task of identifying paddy variety is more difficult and complicated when one is identifying the level of purity of paddy varieties (Liu *et al.*, 2005). In some cases, good quality paddy seed cultivars can be falsified by poor quality cultivars or other fallacious cultivars, affecting the quality, yield and value of rice (Wenwen *et al.*, 2013). Also before releasing a variety, it should be tested for its identity and unity. On the other hand, as a seasonal crop, the harvest of the paddy is being handled in large scale storage warehouses as bulk, together with their representative samples. In case of proper handling of a large number of samples in various events, there is a need for a quick and reliable method for identifying the exact variety or

* Corresponding author (wikyasintha@gmail.com; https://orcid.org/0000-0001-8888-5410)

the cultivar of the varieties, as commercial value, genetic characteristics, and quality of rice mainly depends on the type and grade of the rice variety. Different paddy varieties vary in size, shape, and constitution, and variety is considered as an imperative factor which influences the cooking and processing quality of rice (Namaporn *et al.*, 2011). It is clear that quality parameters of a rice variety have significant impact on consumer demand.

Traditional methods of paddy variety identification which require field planting, such as seedling methods (Sasaki, 2004), morphological characteristics (Hussain *et al.*, 2010) and field planting (Song *et al.*, 2010) need long periods of time. Currently, the variety identification of paddy seed is mainly done by chemical methods such as electrophoresis (Bertrand *et al.*, 2006), HPLC (high performance liquid chromatography), (Huebner *et al.*, 1991) or GC-MS (gas chromatography-mass spectrometry) (Wu *et al.*, 2012). Varietal characteristics are determined through these tests based on proteins, fats, moisture, and amino acids of single seed, and serious overlaps of these components exist among seeds. Further, DNA tests should be performed to determine the structure and composition of genetic materials of seeds (Jing *et al.*, 2019). Although these methods can give relatively more exact results, they have many limitations. Implementations of the chemical methods are time consuming and have shortcomings when the available amount of sample is limited. It is also highly expensive for inspection as well as instrument maintenance. Variety identification through electrophoresis is not applicable to routine control (Bertrand *et al.*, 2006). The measurements require experimental equipment that is complicated, expensive, and highly specialized. Further, non-destructive identification of paddy varieties on a large scale cannot be accomplished by chemical methods or by the paddy field method (Liu *et al.*, 2005).

All the above mentioned methods cannot be used effectively to detect paddy varieties in large scale handling. Instead, it is important to develop a simple, rapid and reliable identification method. (Jing *et al.*, 2019). Therefore, this effort was taken to develop a non-destructive, rapid, and reliable method to identify paddy seed varieties. There are reports on using NIR imaging based paddy seed identification with expensive Indium Gallium Arsenide laboratory setups (Wenwen *et al.*, 2013). The system consists of a laboratory where the sample is kept in a dark room and samples are illuminated using two tungsten halogen lamps (150 W X2). The used wavelength range, 1100-2500 nm, is not usable for simplified field level instrument application. Using lab instruments outfitted with InGaAs (Indium Gallium Arsenide) detectors that span the extended short-wave NIR band (1100–2500 nm), prior research has documented NIR imaging–based rice identification. Although InGaAs instruments are more accurate and sensitive, their high capital cost and the need of specific high power consumed lighting such as tungsten-halogen lamps and dark-room settings, make them inconvenient for field-level robust use. Despite the availability of specific novel portable devices with InGaAs detectors, they are still too expensive for frequent large-scale robust seed handling when it is compared to low-cost, simplified short wave NIR field spectrometers. Moreover, when the 1100–2500 nm range is to be applied, it either requires a stable AC mains power connection or a large battery system to operate the high-energy light sources to stimulate the InGaAs detectors. This not only increases the cost but also demands stable laboratory conditions to avoid strong water absorption interferences. These practical limitations, *i.e.*, cost, energy requirements, and sensitivity to background water absorption have made this wavelength region altogether unsuitable for routine field-level identification of rice seeds. By contrast, the visible–shortwave NIR range (588–1091 nm) is compatible with more affordable silicon-based detectors, enabling the development of lightweight handheld devices.

short wave-near infrared (NIR) spectroscopy has been used for a few decades to investigate physicochemical properties of paddy, such as moisture content (Kawamura *et al.*, 2006), protein content (Barton *et al.*, 1998), amylase content (Delwiche *et al.*, 1995; Barton *et al.*, 1998; Delwiche *et al.*, 1996), whiteness (Barton *et al.*, 1998; Delwiche *et al.*, 1995), milling degree (Delwiche *et al.*, 1996), hardness (Awanthi *et al.*, 2019) and cooking quality (Windham *et al.*, 1997). Beyond research purposes, rapid identification of rice varieties has important practical applications, such as preventing admixture of seed lots in warehouses, maintaining pure parental lines for multiplication trials in seed paddy production for specific agro-climatic zones, supporting seed certification and quality assurance, enabling fair trade, and protection of farmers against mislabeling.

However, there are no reports on using short wave-NIR spectroscopy for the identification of paddy seed varieties as a quick, non-destructive, and chemical free method. Therefore, this study was conducted to assess the possibility of use of SW-NIR spectroscopic methodology to identify paddy seed samples, as it will be a rapid, non-destructive, chemical free, and affordable field level application, and to build up and optimize a spectroscopic SIMCA discrimination model to predict seed samples from five different paddy varieties, as well as to investigate the effect of moisture content in seed samples for NIR prediction results.

MATERIALS AND METHODS

Sample Preparation

A number of 750 seed samples from each of five improved paddy varieties which are recommended for general planting by the Department of Agriculture were obtained from Rice Research Institute (RRI) of Batalagoda, Sri Lanka. The selected varieties, Bg300, Bg352, Bg357, Bg359, and Bw363, are widely planted and improved paddy cultivars which had been produced in the same year to avoid any effect of seed age. Paddy samples were prepared with different moisture contents in order to develop a NIR model which can be used to detect paddy varieties in different stages of post-handling activities such as harvesting, milling and storage. Within the post-harvest process of paddy handling, proper moisture content of the seeds should be available. When harvesting, rice seed moisture content should be within the range20-25%, as higher moisture content results in more losses due to many green and unfilled grains, while lower moisture content results in more losses from shattering. The desired moisture range for paddy threshing is 20-25% in order to minimize incomplete threshing and grain damage. The ideal moisture range for paddy milling is 14-15%. The moisture content of paddy grain should be less than 14%, and the ideal is 12% for long time storage (IRRI, 2021). Therefore, paddy samples, with moisture contents of 23%, 15% and 12% were prepared to acquire spectral data. In this study, systematic laboratory control was applied to maintain paddy seed moisture content at 23%, 15%, and 12%. The final MC% was assessed using the time-consuming gravimetric method. However, in practical applications, this procedure can be simplified. For instance, according to the literature, the same NIR spectrometer can be employed to detect paddy seed moisture content almost instantly and with high accuracy by applying a PLS model (partial least squares) developed for this purpose. In the way forward, quick moisture standardization could also be achieved by creating specific RH environments, rather than soaking with various methods, through the use of saturated solutions of a salt such as lithium chloride, in a closed container.

Preparation of raw paddy samples

Paddy samples were winnowed to remove paddy straw and other impurities. Then all the samples which were in storage condition were weighed, and the moisture content of samples was determined through the oven drying method (IRRI, 2021). A spectrum was taken from each of the 150 seeds of each variety.

Preparation of paddy samples under different moisture contents

To analyze model performances based on the various moisture contents, seeds were soaked in cold water for 24 hrs. Then the paddy seed samples were dehydrated until the moisture content reaches 23% and the samples were packed using polysack bags. Then they were kept under room temperature (27 °C) in order to allow reaching equilibrium moisture among all paddy seeds. All rice samples were treated in the same manner—soaking, draining, drying, and packing into polysack bags. Each sample reached a moisture content of 23% as confirmed by the gravimetric method (IRRI, 2021). Then spectra were taken from each sample. The same procedure was repeated to prepare paddy samples with 15% and 12% moisture content.

Data collection

All individual 750 paddy samples were used to obtain the NIR spectra to be analyzed and to use for the model building process.

Instrument and Spectra Acquisition

A handheld type NIR Spectrometer FQA-NIR Gun (588-1091 Shizuoka Shibuya Seiki, Hamamatsu, Japan), was used to acquire the reflectance spectra from 588 nm up to 1091 nm at 2 nm steps of five paddy seed varieties. In order to improve the accuracy of spectral data, background noises had to be avoided. The rigid steel probe in the original instrument [1(a)] was not suitable for paddy seed spectra measurement. A newly modified, specially designed best fitted probe [Figure 1(b)], which was reported previously and tested for biological samples, was used to avoid environmental interferences (Balasuriya *et al.*, 2010). A custom-built probe extension was fabricated using a foamed rubber block measuring 50 × 40 × 50 mm (length, width, and height). The original NIR instrument probe (10 × 15 × 15 mm) was inserted into the top surface of the block by removing a cube of matching size. Two cylindrical holes, each 10 mm in diameter, were drilled along the 50 mm height of the rubber block to provide a path for the light source to reach the sample and for the light of the reflectance spectra to return to the instrument detector. After doing several trials, a white polypropylene container of 5 cm diameter was selected to hold paddy seeds at the time of taking spectra, as it avoids the light interference further. The cylindrical sample cups measured 5 cm in diameter and 10 mm in height. Each cup was filled to its full 10 mm height, and a quartz screen was positioned on top during measurements. The bottom surface of the probe extension was gently pressed against the quartz screen while spectra were recorded, ensuring that no gap existed between the probe extension and the quartz surface above the sample cup.

This approach is intended for the qualitative discrimination of paddy samples rather than for the detection of material quantities based on their absolute chemical constituents. For example, in quantitative mode prediction using a PLS model, or when identifying the spectroscopic signature of paddy sample spectra, the influence of the custom probe extension or the sample holder material would be a critical consideration.

The method presented here is based on supervised classification, utilizing known paddy seed samples as the reference class to ensure accurate identification. As the applied methodology has successfully differentiated the samples with satisfactory results, consistent with the reference samples, the effect of the sample holder material on the spectra becomes negligible. Nevertheless, we acknowledge that the sample holder and probe extension material may potentially influence the information captured by the spectra. However, since these factors are common to all spectra collected in this study, any such influence is effectively cancelled out, rendering their impact on model accuracy insignificant.

Seed samples were filled into 5 cm diameter container and the seed surface was closed by the quartz screen when the spectra were taken [Figure 1(c)]. The sample cup was always filled with seeds to its top edge, ensuring that an equal amount of paddy was exposed in every spectral observation. This arrangement in the sample presentation setup eliminated any possible interference from ambient light. Furthermore, no measurable temperature increase was detected during the measurements, as the spectral integration time was as short as 100 milliseconds. One hundred and fifty seed samples per variety were illuminated by the NIR spectrometer and the reflectance spectra were recorded. Accordingly, there were 750 spectra in the data set.

(a) (b) (c)

Figure 1: NIR spectra acquisition (a - The NIR instrument with original rigid steel probe. b - The instrument with the newly modified best fitted probe. c - Aequiring the NIR spectra using the modified probe)

Spectral data analysis

Prior to data analysis, the head and tail of spectral data that contained noise and baseline shift were removed. Although the instrument covered 588 – 1091 nm, noisy edge regions below 630 nm and above 1080 nm were clipped during preprocessing, leaving an effective working range of 630–1080 nm for analysis. Then principal component analysis (PCA) was done in order to remove outliers of the data set according to the Mahalanobis distance. From the total of 750 spectra, 20 spectra (≈ 2.6%) were excluded. These were likely due to physical breakdown of paddy (*e.g.*, cracked seed coats or broken seeds) or interference from outside light caused by insufficient pressure applied when pressing the probe against the quartz screen. To minimize outliers, precautions such as careful seed placement, applying gentle but consistent pressure of the probe against the quartz screen, and repeating doubtful scans were adopted.

Out of the 750 spectra, 500 spectra were used for the training set and 250 spectra were used to create a prediction set for the model evaluation process. Spectra were selected alternatively as 10 spectra for training set and 5 spectra for prediction set. Paddy seed variety prediction was achieved by soft independent modeling of class analogy (SIMCA). Spectra from each variety Bg300, Bg352, Bg357, Bg359, and Bw363 were assigned as discriminate class 1, 2, 3, 4, and 5 respectively in the SIMCA algorithm in Pirouette software (Informetrix, Woodinville, WA).

Moisture content (MC) has a prominent effect on near-infrared (NIR) spectra because water strongly absorbs wavelengths around 746 nm, 970 nm etc., which correspond to the third overtone of the water absorbance bands. Increasing MC therefore can alter the baseline, peak intensity, and scattering properties of the spectra, making different mathematical treatments effective for the samples having different moisture contents.

RESULTS AND DISCUSSION

Prediction performance of SIMCA

The NIR spectra obtained from paddy samples were used to prepare a SIMCA model based on different mathematical treatment options. The best prediction results obtained for the tested paddy varieties under different moisture contents were summarized as follows. There were 150 sample spectra from one variety, of which 100 were used to prepare the model set and the remaining 50 were used as the prediction set. Out of the 50 samples, the number of spectra correctly classified was expressed as a percentage in Table 1. For instance, 43/50 corresponded to 86%. In some cases, a few outliers in the prediction set produced slightly different results, such as 46/49 = 94%. These findings highlight that while the model performed consistently well, occasional misclassifications can influence the overall prediction accuracy.

As shown in the table, there is an increasing trend in the results with the elevated moisture in the samples. from the table, it is clear that in the prediction models, the maximum number of erroneously classified spectra was recorded as 10 out of the 50 spectra (prediction under 10% and 12% moisture).

Table 1: Summary of results under different moisture contents

	10% MC.	12% MC.	15% MC	23% MC	% prediction
Bg 300	84	84	94	95.91	89.5
Bg 352	84	82	80	86	83
Bg 357	80	80	90	100	87.5
Bg 359	80	86	92	92	87.5
Bg 363	82	81	83	87	83.25
% prediction	82	82.6	87.8	92.182	

SIMCA Prediction results of 10% moisture content

The highest number of correct classifications were obtained based on the mathematical treatments given below, for moisture content of 10%. Fifty spectra were evaluated for each variety, and the highest number of misclassifications (10 spectra) were recorded for the Bg357 and Bg359 varieties.

Mathematical treatment
- » Baseline correction (linear fit)
- » Mean centre (Optimal factors – 10)
- » 1^{st} derivative 9
- » Smooth

According to the above model configuration, 80% or higher correct classifications were obtained for five paddy seed varieties with 10% moisture content. It was observed that the maximum accuracy rate of prediction was possible only up to 84%.

Table 2: Summary of results for 10% moisture content

Variety	% prediction
Bg 300	84
Bg 352	84
Bg 357	80
Bg 359	80
Bg 363	82

SIMCA Prediction results of 12% moisture content

At 12% moisture content, more than 80% correct classifications were recorded for all the varieties based on the mathematical treatments given below. Only one sample of no match was recorded for the variety Bg352.

- Mean centre10
- 1^{st} derivative 7
- Smooth
- Baseline correction (linear fit)

Table 3: Summary of results for 12% moisture content

Variety	% prediction
Bg 300	84
Bg 352	82
Bg 357	80
Bg 359	86
Bg 363	81

SIMCA Prediction results of 15% moisture content

According to the SIMCA performance of 15% moisture content, 80% correct classifications for the variety Bg352 and 83.33% for the variety Bg363 were recorded, and more than 90% correct classifications were obtained for varieties Bg300, Bg357, and Bg359.
Mathematical treatment

- Mean center10
- 1^{st} derivative5

Table 4: Summary of results under 15% moisture content

Variety	% prediction
Bg 300	94
Bg 352	80
Bg 357	90
Bg 359	92
Bg 363	83

SIMCA Prediction results of 23% moisture content

Based on pereto 5 and 2^{nd} derivative 15 mathematical treatment, the highest correct classification results of 100%, was recorded for the variety of Bg357 and more than 85% correct classifications were recorded for the rest of the varieties as well.

Table 5: Summary of results under 23% moisture content

Variety	% prediction
Bg 300	95.91
Bg 352	86
Bg 357	100
Bg 359	92
Bg 363	87

According to the prediction results, different varieties have shown different correct classifications rates in the SIMCA models constructed for different moisture contents. However, more than 80% correct classifications were obtained for all selected varieties under tested moisture content. The sample arrangement in the SIMCA model configurations shows that the separation of different varieties in three-dimensional principal component space was arranged according to the specific characteristics of each variety. According to the previously reported research (Li *et al.*, 2008) on "Non-destructive prediction of storage time of paddy seeds by Vis/NIR spectroscopy", the best model was achieved with a high discrimination accuracy rate of 97.5%. Thus, NIR technology has proved consistency in its successes for various applications in the paddy industry.

Spectra of misclassified samples showed weaker absorption in the 970 nm and 1030 nm regions, corresponding to O–H stretching overtones of water and C–H vibrations of starch. Since the same varieties may differ slightly in starch (80–84%) and protein (7–9%) content (Xinkang *et al.*, 2023), these compositional differences are likely to modulate the observed bands. The results suggest that both starch structure and protein secondary vibrations contribute to varietal discrimination, with subtle variations being sufficient to influence classification outcomes.

Higher model performances, more than 86% of correct classifications, were obtained at 23% moisture out of the tested moisture contents of 10%, 12%, 15% and 23%. The amount of moisture should be less than 14%, and preferably less than 12% for extended storage times (Gummert, 2010). A moisture amount of 10% is the best moisture level for the storage of rice seeds. Hence, the SIMCA model developed for the 10% moisture level can be used to identify tested rice varieties in storage condition.

The highest correct classifications are recorded at a moisture content of 23%. In view of the SIMCA performances for 10%, 12%, 15%, and 23% moisture contents, it is noted that, when increasing the moisture content of paddy seeds, the number of correct classifications is also increasing. Some paddy samples are recorded as "no matches". The "no matches" could potentially be reduced by increasing the number of spectra in the model-building set (Balasuriya *et al.*, 2010).

The tested five varieties of paddy are very similar in physical characteristics such as their colour, shape and kernel characters. So the spectral features of all varieties are almost similar. One of the research projects, similar to our work, on "transmission NIR spectroscopy for discrimination of geographical origins of rice" by Jaejin *et al.*, (2010) has indicated that the overall spectral features between Korean and Chinese rice samples which were used for their research were almost identical. Rice is composed of approximately 80~84% starch and 7~9% protein. Fats, minerals, and vitamins are the other minor constituents. Therefore, the resulting NIR features are obviously dominated by those of starch. As starch is a huge polymer of glucose, with a molecular weight ranging from 50,000 to 200,000 Daltons, it would be hard to find spectral differences from starch for discrimination. Conversely, although the protein content is relatively much lower, this could provide valuable spectral clues for discrimination, because of the higher variability in the chemical structure (Xia *et al.*, 2013). Therefore, the latent variable working behind the classifications would be these differences in the varieties. Other than that, the micro level structural variations and colour variations that are invisible to the human eye in the seed coat may play an important role. The limited financial resources available for this project did not permit detailed constituent analysis to be included in this study. Nevertheless, a research proposal is currently being formulated to enable more comprehensive investigations into the biochemical and structural bases of correct varietal identification. Such an extension would provide stronger evidence for the mechanistic role of starch and protein features in spectral classification.

Average spectra observation of paddy seed varieties

According to the average spectra, samples of paddy seed varieties showed relatively higher NIR radiation reflectance at 10% moisture content and lowest reflectance at 23% moisture content. That means higher light absorbance in higher moisture content and lower light absorbance in lower moisture content. Similar to our findings, Sineenart & Sontisuk (2012) have indicated in their research on "Detection for moisture content of sweet tamarind flesh by transmittance short wavelength near-infrared spectroscopy", that averaged absorbance of low moisture content was reported less than that of middle and high moisture content. However, the pattern of light absorbance has altered with 12% and 15% moisture contents.

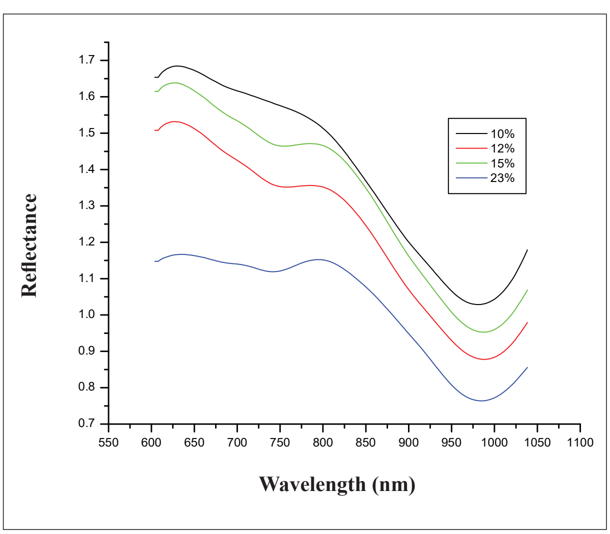

Figure 2: Average spectra of Bg 300

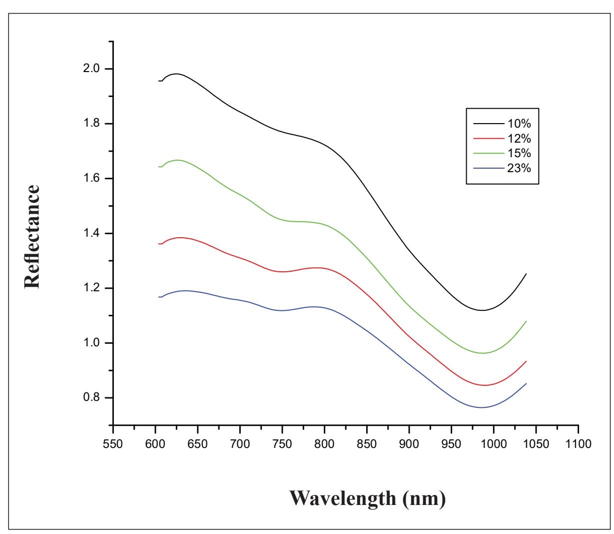

Figure 3: Average spectra of Bg 352

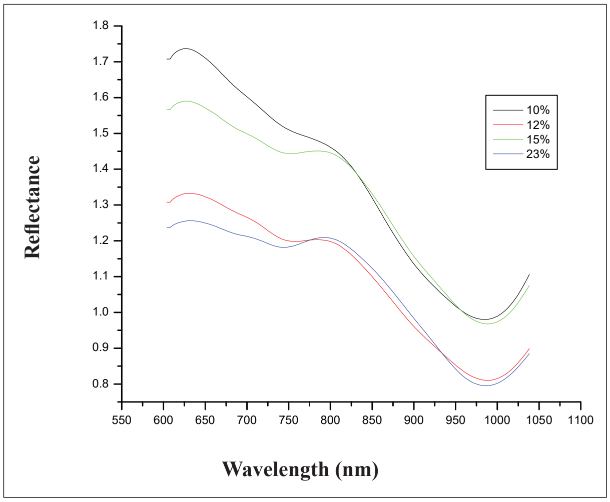

Figure 4: Average spectra of Bg 357

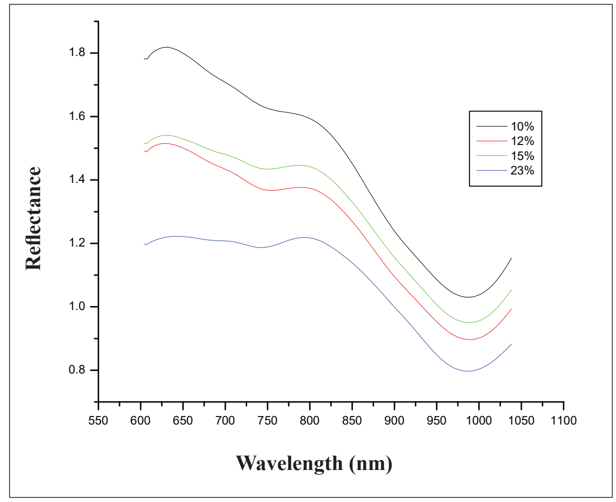

Figure 5: Average spectra of Bg 359

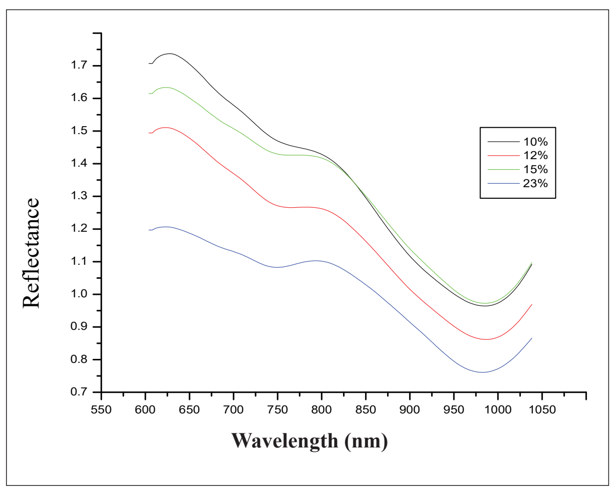

Figure 6: Average spectra of Bw 363

A considerable effort was made to illustrate and present the spectral differences between samples under different scenarios. However, due to the very low amplitude on a single y-axis scale, it was not possible to present these differences graphically. This limitation underscores the sensitivity of NIR methods, where subtle but meaningful differences may not always be visually discernible yet can still be captured effectively through chemometric modeling.

Different rice varieties at the same moisture content have identical relative moisture on a dry matter basis, according to the gravimetric definition. However, the absolute amount of water present in each sample cup may vary across varieties because of differences in bulk density arising from kernel size and shape. This observation highlights the importance of considering physical grain characteristics in addition to chemical composition when interpreting NIR-based varietal discrimination. When looking at the average spectra of all paddy varieties it is clear that the pattern of absorbance at different amounts of moistures is almost similar from 630 to 800 nm. That means, macro level individual characteristics of all the samples are more or less similar at the visible region. When considering average spectra, their pattern of absorbance was slightly differing when the moisture contents in the samples were differing at the NIR range 800 nm to 1100 nm, compared to the visible region. According to the distance between baselines of different moisture levels, it is clear that when the prevailing moisture content of the samples is changed, the amount of radiation absorbance of the varieties also change.

There is a strong relationship between water and biological samples and their functionality. Aquaphotomics is the new discipline that is used to define the application of spectrophotometry in the visible/near infrared region to the understanding of the effect of water on the structure and functioning of biological systems (Tsenkova, 2007). This result are consistent with the previous report that all of these sample types are dominated by water and water strongly absorbs NIR light energy. This research has demonstrated the consistency of the concept Aquaphotomics that higher moisture in the sample results in higher radiation absorbance and better accuracy in results.

Discriminating power

The discriminating power graphs in the SIMCA sub plot of results indicate the most effective wavelength for the classification. According to the graph wavelengths close to 740 nm have higher discriminating power. Therefore, we may further investigate possible simplified instrumentation for paddy seed variety detection by using these wavelengths.

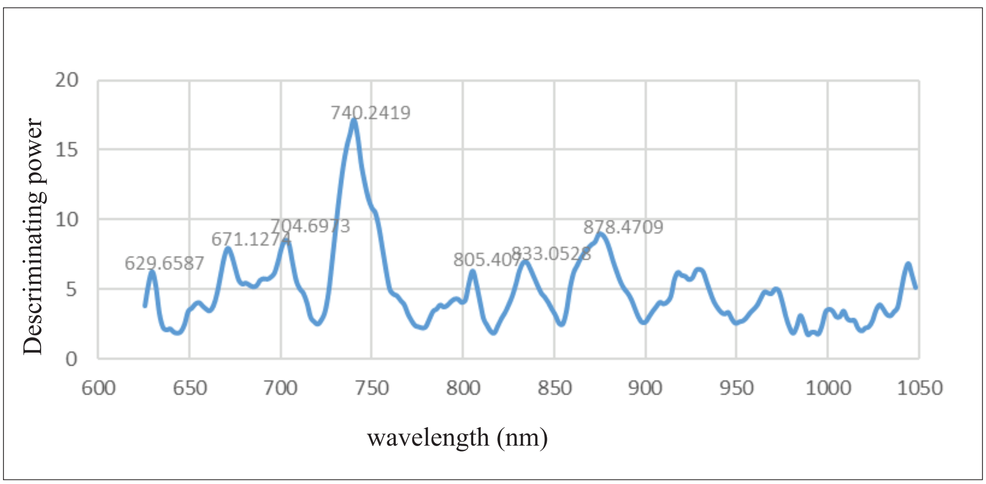

Figure 7: SIMCA discriminating power

CONCLUSIONS

An increasing trend in correct classification accuracy was observed with elevated sample moisture levels, reaching the highest average accuracy (92%) at 23% MC. This can be attributed to the fact that water molecules exhibit strong and distinct absorbance bands in the NIR region for the biological sample (Jinendra *et al.*, 2010), particularly around 740 nm and 878 nm. At higher MC levels, these water-associated signals enhance overall spectral contrast, making varietal differences more pronounced and easier for the SIMCA model to capture.

The mis-classification of up to 20% of spectra had little impact on the overall identification of paddy seed varieties. For instance, when spectra of 50 samples were used to represent a single variety in the prediction set, then a 20% error means that only 10 spectra were misclassified while the remaining 40 were correctly assigned into their correct group. Since the majority of spectra clustered within the correct varietal group, the intended variety is reliably revealed. This demonstrates that even with moderate error rates, replicate sampling provides robustness and ensures accurate varietal classification.

The developed SIMCA models for the identification of investigated studied rice varieties, namely Bg300, Bg352, Bg357, Bg359, and Bw363, were successfully accomplished at a rate of 80% or better accuracy. The developed methodology provides higher identification accuracy when multiple spectra from a single sample are analyzed. By collecting several spectra and assigning the sample to the class indicated by the majority of correctly classified spectra, the overall reliability of the sample identification is enhanced. Elevated moisture content in the paddy sample indicated higher NIR absorbance and also higher accuracy. The research successfully demonstrated the potential of using NIR spectroscopic SIMCA approach for the identification of five tested paddy seed varieties under 10% to 23% moisture range.

Acknowledgement

We gratefully acknowledge Shibuya Seiki Co Ltd., Japan, for providing the NIR spectrometer to carry out this research and Dr.(Mr.) Lionel Nugaliyadde for helping to reach the Rice Research and Development Institute (RRDI), Bathalagoda, Sri Lanka, to get the seed samples from studied rice varieties.

REFERENCES

Xinkang, L., Chunmin, G., Lin, W., Liting, J., Xiangjin, F., Qinlu, L., *et al.* (2023). Rice Storage Proteins: Focus on Composition, Distribution, Genetic Improvement and Effects on Rice Quality. *Rice Science*, 207-221. doi.org/10.1016/j.rsci.2023.03.005

Awanthi, M., Jinendra, B., Nawaratne, S., & Nawaratne, C. (2019). Adaptation of visible and short wave Near Infrared (VIS-SW-NIR) COMMON PLS model for quantifying paddy hardness. *Journal of Cereal Science, 89*, 102795. doi:org/10.1016/j.jcs.2019.102795

Balasuriya, J., Katsutomo, T., Shinichiro, K., Maria, V., Shinya, Y., & Roimiana, T. (2010). Near infrared spectroscopy and aquaphotomics: Novel approach for rapid *in vivo* diagnosis of virus infected soybean. *Biochemical and Biophysical Research Communications, 397*, 685-690.

Barton, F., Windham, W., Champagene, E., & Lyon, B. (1998). Optimal Geo-metries for the Development of Rice Quality Spectroscopic Chemometric Models. *Cereal Chemistry, 75*(3), 315-319.

Bertrand, D., Robert, P., & Loisel, W. (2006). Identification of some wheat varieties by near infrared reflectance spectroscopy. *Journal of the science of food and Agriculture spectroscopy, 36*(11), 1120–1124.

Cheng, F., & Ying, Y. (2004). Machine vision inspection of rice seed based on Hough transform. *Journal of Zhejiang University Science, 5*, 663- 667.

Delwiche, S., Bean, M., Miller, R., Webb, B. D., & Williams, P. (1995). Apparent Amylose Content of Milled Rice by Near-infrared Reflectance Spectrophotometry. *Cereal Chemistry, 72*(2), 182-187.

Delwiche, S., McKenzie, K., & Webb, B. (1996). Quality Characteristics in Rice by Near-infrared Reflectance Analysis of Whole-grain Milled Samples. *Cereal Chemistry, 73*(2), 257-263.

DOA. (2015). *Rice.* Retrieved 10 21, 2015, from Department of Agriculture: http://agridept.gov.lk/index.php/en/journals

Gummert, M. (2010). *Seed quality.* Retrieved 05 22, 2015, from knowledgebank: http://www.knowledgebank.irri.org/step-by-step-production/pre-planting/seed-quality

Guzman, J., & Peralta, E. (2008). Classification of Philippine rice grains using machine vision and artificial neural networks. *World conference on agricultural information and IT* (pp. 41-48.). Tokyo, Japan : IAALD AFITA WCCA.

Huebner, F., Hussain, A., Lookhart, G., Bietz, J., Bushuk, W., & Juliano, B. (1991). Discrimination of sister-line IR rice varieties by polyacrylamide-gel electrophoresis and reversed-phase high-performance liquid-chromatography. *Cereal Chemistry, 68*, 583-588.

Hussain, Z. P., Azmi, M., & Ahmad, S. O. (2010). Morphological Study of the Relationships between Weedy Rice Accessions (*Oryza sativa* Complex) and Commercial Rice Varieties in Pulau Pinang Rice Granary Area. *Tropical Lifesciences Research, 21*(2), 27-40.

IRRI. (2021). *Measuring Moisture Content*. Retrieved 08 12, 2021, from http://www.knowledgebank.irri.org/step-by-step-production/postharvest/milling/milling-and-quality/measuring-moisture-content-in-milling#in-drying

Jaejin, K., Jihye, Y., Si Jun, P., & Hoeil, C. (2010, 10 15-18). Transmission NIR Spectroscopy for Discrimination of Geographical Origins of Rice. *The 2nd Asian Symposium on Near Infrared Spectroscopy*, p. 121.

Jing, Z., Mingliang, L., Tao, P., Lijun, Y., & Jiemei, C. (2019). Purity analysis of multi-grain rice seeds with non-destructive visible and near infrared spectroscopy. *Computers and Electronics in Agriculture, 104882*, 164.

Kawamura, S., Natsuga, M., & Itoh, K. (2006). Visual and Near-infrared Reflectance Spectroscopy for Determining Physicochemical Properties of Rice. *American Society of Association Executives Annual International Meeting* (pp. 1069- 1076). Orlando: American Society of Agricultural and Biological

Engineers.

Li, X., He, Y., & Wu, C., (2008). Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. *Journal of Stored Products Research, 44*(3), 264–268.

Liu, Z., Cheng, F., Ying, Y., & Rao, X. (2005). Identification of rice seed varieties using neural network. *Journal of Zhejiang University Science , 6*(11), 1095- 1100.

Namaporn, A., Sumaporn, K., & Athapol, N. (2011). Rapid Variety Identification of Pure Rough Rice. *Cereal Chemistry, 88*(5), 490-496. doi:10.1094/CCHEM-03-11-0025

Sasaki, R. (2004). Characteristics and Seedling Establishment of Rice Nursing Seedlings. *Japan Agricultural Research Quarterly, 38*(1), 7-13. doi:10.6090/jarq.38.7

Song, C., Weihui, Z., Fengrong, Z., & Guoping, Z. (2010). Effect of Planting Method on Grain Quality and Nutrient Utilization for No-Tillage Rice. *Communications in Soil Science and Plant Analysis, 42*(11), 1324-1335. doi:.org/10.1080/00103624.2011.571737

Tsenkova, R. (2007). AquaPhotomics: water absorbance pattern as a biological marker for diesease diagnosis and diesease undestanding. *Aquaphotomics, 18*, 15-17.

Wenwen, K., Chu, Z., Fei, L., Pengcheng, N., & Yong, H. (2013). Rice Seed Cultivar Identification Using Near-Infrared Hyper spectral Imaging and Multivariate Data Analysis. *Journal of sensors, 13*(7), 8916-8927.

Windham, W., Lyon, B., Champagne, A., Barton, E., Webb, B., McClung, F., *et al*. (1997). Prediction of Cooked Rice Texture Quality Using Near-infrared Reflectance Analysis of Whole-grain Milled Samples. *Cereal Chemistry, 74*(5), 626- 632.

Wu, J., Yu, H., Dai, H., Mei, W., Huang, X., Zhu, S., & Peng, M. (2012). Metabolite profiles of rice cultivars containing bacterial blight-resistant genes are distinctive from susceptible rice. *44*, 650- 659.

Xia, L., Shen, S., Liu, Z., & Sun, H. (2013). Identification of geographical origins of rice with pattern recognition technique by near infrared spectroscopy. *Spectroscopy and Spectral Analysis, 33*, 102- 105.