ADVANCES IN AGRICULTURAL AND FOOD RESEARCH JOURNAL

Short Communication

Characterisation of Weedy Rice Seeds using Principal Component Analysis

Rashidah Ruslan[1,2], Siti Khairunniza Bejo[1,3,4]*, Mahirah Jahari[1,3], Ibni Hajar Rukunuddin

[1]Department of Biological and Agricultural Engineering, University Putra Malaysia, Serdang, Selangor 43400, Malaysia, jmahirah@upm.edu.my

[2]Faculty of Chemical Engineering Technology, University Malaysia Perlis, Arau 02600, Perlis, Malaysia, rashidahruslan@unimap.edu.my

[3]SMART Farming Technology Research Center, Faculty of Engineering, University Putra Malaysia, Serdang, Selangor 43400, Malaysia, skbejo@upm.edu.my

[4]Institute of Plantation Studies, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia, skbejo@upm.edu.my

*Corresponding author: Siti Khairunniza Bejo, Department of Biological and Agricultural Engineering, University Putra Malaysia, Serdang, Selangor 43400, Malaysia; skbejo@upm.edu.my

Abstract: The weedy rice contamination in certified rice seed samples greatly impacts the Malaysian rice seed industry. The existing manual process to identify the weedy rice seed is only based on the physical appearance of the seed. The physical characteristics such as morphology, colour and texture image of the seeds were captured and analysed using image processing and the application of machine vision to understand the physical characteristics of the weedy rice seed. The objective of this study is to understand the physical characteristics of the weedy rice seed using Principal Component Analysis (PCA) transformation. A total of 7350 images of cultivated rice seeds from five major varieties and 895 images of weedy rice seeds were acquired using machine vision setup, and 67 features from the three major parameters (morphology, colour, and texture) were extracted. The test of equality of means based on Wilks' Lambda was performed to assess significant differences among the group parameters. PCA transformed data into principal components. The relationships between weedy rice seed and cultivated rice seed samples were examined through the score and loading plots of the PCA analysis. This newly transformed data visualises the experimental data's underlying structure and helps identify the parameters distinguishing between weedy rice and cultivated rice seed. The results on the PCA score plot have shown overlapping areas between the cultivated and weedy rice seeds, indicating high similarities between the seed samples.

Keywords: Morphology; multivariate analysis; computer vision; paddy seed

Received: 17th May 2023

Accepted: 4th October 2024

Available Online: 17th June 2025

Published: 1st July 2025

Citation: Ruslan, R., Bejo, S. K., Jahari, M., *et al*. Characterisation of weedy rice seeds using principal component analysis. *Adv Agri Food Res J* 2025; 6(2): a0000534. https://doi.org/10.36877/aafrj.a0000534

1. Introduction

Weedy rice is known as a weed species in a rice field. The weedy rice infestation is not a local problem in Malaysia, but it was reported worldwide and has spread to many rice-growing fields in countries. The infestation in some countries such as Europe, the USA, Cuba and Senegal were estimated at 30%, 70%, 80% and 50% of the rice field, respectively (Nadir *et al.*, 2017). In China, 15% of rice planting area practices a direct-seeded system, and it was estimated that six million hectares of Northeastern China rice fields had been infested with weedy rice (Sun *et al.*, 2013).

Weedy rice is also known as 'red rice' in other parts of the world, and it is called 'padi angin' in Malaysia due to its easy-shattering traits. Regardless of its name, weedy rice has similar morphological and physiological characteristics to cultivated rice, making it difficult to differentiate the plants (Song *et al.*, 2014). However, the weeds' phenotype characteristics are the easiest and most accessible to farmers to identify weedy rice morphotypes in the rice fields. Many researchers have recorded the weedy rice plant traits. According to Zhao *et al.* (2018), Ahmed *et al.* (2012), and Karim *et al.* (2004) the weedy rice plant is taller than cultivated rice, with shorter grain-filling stages in the early maturity stage. Some of the seed characteristics are easily shattering, strong dormancy (Nadir *et al.*, 2017) and can withstand extreme weather, and the seed grain often has a pigmented aleuronic layer on the endosperm (hull colour). Some weedy rice seeds also have an awn or no awn on the grain.

The weedy rice morphotypes exhibit similar physical traits with the cultivated rice seed in terms of hull colour, pericarp colour and seed shape. Research by Sudianto *et al.* (2016) has classified the weedy rice seed based on two significant grain traits: the hull colour and the awn presence for quick identification. However, the awn presence is sometimes not practical as it is already broken once the seed has undergone cleaning and pre-processing in the seed plant. In terms of size and shape, none of the studies reviewed by Nadir *et al.* (2017) reported the exact weedy rice seed shape and length. Nonetheless, a review by Ruzmi *et al.* (2021) classified the weedy rice seed shape based on grain length/width ratio into slender, bold or round classes where most of the weedy rice seeds are either long/slender or round/short.

It is essential to understand the weedy rice seeds based on quantitative measures such as the morphology, colour and texture based on the seed images. Comparing the quantitative parameters of the weedy rice seed with cultivated rice seed is essential in designing the mechanism to remove weedy rice seed from the certified cultivated rice seed. The wide application of machine vision in the grain industry (Qiu *et al.*, 2018; Huang & Chien, 2017;

Kuo *et al.*, 2016) has proved that the varietal classification of rice seed is possible. The application of multivariate analysis, such as the Principal Component Analysis (PCA) as an initial visualisation tool (Ansari *et al.*, 2021; Aznan *et al.*, 2021) on the descriptive dataset provides insight into how the seed samples relate to the extracted features. It is also used as an investigative step to spot hidden patterns in the dataset (Maione & Barbosa, 2019).

Therefore, this study was formulated to study the relationships between the weedy rice seed and cultivated rice seeds and the relationships of the extracted features to the two seed groups using the PCA method. Machine vision application allows image acquisition of the weedy rice seed and quantitative measures of the seed grain, such as the morphology, colour, and textural features to be extracted.

2. Materials and Methods

2.1 Sample Collection

A total of 7350 cultivated rice seed kernels were acquired from five different varieties, namely MR297, MR220 CL2, MR219, UKMRC2 and UKMRC8 and 895 weedy rice seed kernels were collected from the Department of Agriculture Teluk Cengai, Kedah. The samples of weedy rice seeds represent the variants of weedy rice found in the MUDA area. The weedy rice seed samples were acquired while sampling other seeds/weed seeds for certification under the Paddy Seed Certification Scheme.

2.2 Image Acquisition and Processing

A 6-megapixel resolution of a CMOS area scan camera (MVCA060-10GC, HIK Vision) has a 7.2 mm x 5.4 mm and pixel size of 2.4 µm x 2.4 µm coupled with a C-mount lens type with a focal length, f of 25 mm was set up on a machine vision prototype. The machine vision used a low-angled diffused square front light (DLW2-60-070-1W-24V) and a high-intensity backlight (BHLX3-00-320x320-X-W-24V) for uniform illumination of the seed samples. The combination of the two lighting removes the seed shadows and glare. Backlight illuminated from beneath the seed samples placed on top of the diffused plate of the lighting module. Before image acquisition, the camera settings, such as the iris opening, exposure time, balance ratio selector, and black level, were adjusted and remained constant throughout the sample's imaging to ensure uniformity.

The seed sample images were acquired in RGB colour images. The seed placement was ensured non-touching and no overlapping by placing the seeds in a seed holder. The seed holder holds 15 kernels (5 rows x 3 columns) with a size of 40 mm x 28 mm (W x H) that fit the field of view of the area scan camera to capture 15 seeds per image. Figure 1 shows the sample images captured for this study.

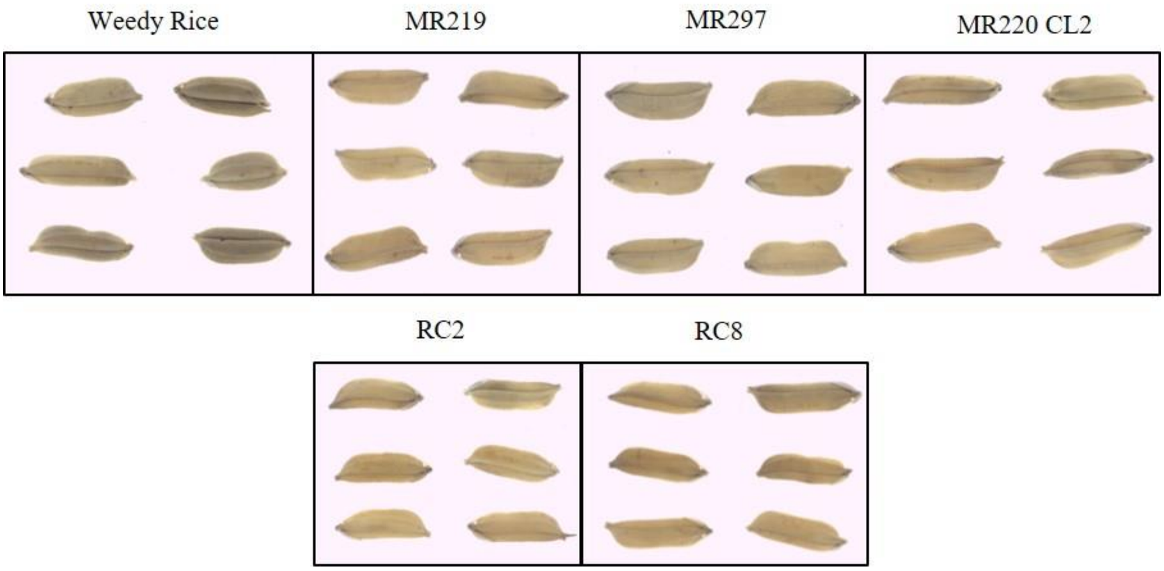

Figure 1. The RGB images of the weedy rice and five cultivated rice seed varieties were captured using the machine vision setup.

The image processing was done in LabVIEW 2016 (National Instruments, Texas, USA). The RGB image was segmented from the background using histogram thresholding (IMAQ Colour Thresholding) by removing the white pixel value above 250 and replacing it with a converted 8-bit greyscale image. The greyscale was converted to a binary image with values 0 as the background and 1 as the sample particle.

2.4 Parameters Extraction

To extract the features of the seed sample, IMAQ Particle Analysis was used to calculate the selected parameter of the features, such as determining the boundary of the particle sample and the centre of the mass as the reference point for morphological characteristics of the sample. Three main parameters were extracted from each seed kernel: the morphology, colour, and texture features for each RGB image. Thirteen morphological, 24 colours and 30 textural features were extracted from the RGB images resulting in 67 features.

The morphology features described the characteristics of each of the seed kernels acquired from digital images features such as the area, perimeter, convex hull perimeter maximum Feret diameter, major axis length, minor axis length aspect ratio, thinness ratio, ellipse ratio, hydraulic radius, angle orientation, moment X and moment Y. The colour features described the seed colours' mean, standard deviation, the minimum and maximum value of the Red (R), Green (G), Blue (B), Hue (H), Saturation (S), and Value (V) planes. The textural features were extracted from the grey level co-occurrence matrix (GLCM), which characterises the textures of an image using statistical measures such as the mean, variance, uniformity, entropy, maximum probability, correlation, homogeneity, contrast, cluster shade and cluster prominence for each of the colour plane (R, G, B and H, S, V).

2.5 Multivariate Analysis

The test of equality of means was conducted under the Stepwise Discriminant Analyses (SDA) using IBM® SPSS® Statistics software version 25 (New York, USA). The extracted features' descriptive means and univariate analysis were based on Wilk's Lambda value in the Stepwise Discriminant Analysis.

The PCA was performed using The Unscrambler® X (ver-10.4) (Oslo, Norway) from the CAMO software. The dataset was normalised before the PCA analysis. PCA projection method visualised the information on orthogonal principal components (PC) derived from the variables and samples in the dataset (Maione & Barbosa, 2019). The investigation of the score and loadings of the PC was an initial step before classification modelling. Each PC were ranked, and the first and second PC always carried the most information than any of the following PC. The transformed dataset on the PC axes was uncorrelated (Ansari *et al.*, 2021). In this analysis, the score plot of the PCA analysis was used to identify the similarities or differences between the weedy rice seeds and cultivated rice seed samples. The closer the samples, the higher the similarities within the score plot. The observation on the loading plot describes the relationship between the parameters extracted from the seed kernels.

3. Results

3.1 Test of Equality of Means

Table 1 shows the first 15 features ranked by the Wilks' Lambda under the test of equality of means. The test of equality of means has demonstrated that all the features are significantly different (SD) with an SD value less than 0.05. According to the full results, all components are significantly different showing that all variables have the same potential to be included on the classification model. The first three lowest Wilk's Lambda were Std Dev R and Std Dev V (0.557), followed by the M variance value (0.593). The variables with low Wilks' Lambda values suggest that they were better at distinguishing between groups of variables. Further, the lower Wilks' Lambda indicates they significantly contribute to the stepwise analysis (Abdullah *et al.,* 2001).

Table 1. Summary of the first 15 features ranked by Wilks' Lambda on the test of equality means.

No.	Variable	Wilks' Lambda	F-Statistic	df1	df2	SD
1.	Std Dev R	0.557	1307.201	5	8219	0.000
2.	Std Dev V	0.557	1306.333	5	8219	0.000
3.	M variance value	0.593	1129.222	5	8219	0.000
4.	M standard deviation value	0.623	992.702	5	8219	0.000
5.	Mean V	0.631	960.174	5	8219	0.000
6.	Mean R	0.632	958.980	5	8219	0.000
7.	Std Dev G	0.638	934.587	5	8219	0.000
8.	M Mean Value	0.641	918.837	5	8219	0.000
9.	M SeedMaxFeretDiameter	0.642	917.431	5	8219	0.000
10.	SeedMaxFeretDiameter	0.642	917.158	5	8219	0.000
11.	M ConvexHullPeri	0.652	876.623	5	8219	0.000
12.	ConvexHullPeri	0.653	874.254	5	8219	0.000
13.	Contrast B	0.657	857.266	5	8219	0.000
14.	Contrast G	0.658	853.795	5	8219	0.000
15.	Contrast R	0.661	841.745	5	8219	0.000

3.2 PCA

Figure 2 shows the score plot for the cultivated rice and weedy rice seeds and the loadings plot of the RGB parameter of the morphology, colour and texture. The PC axes used were PC1 and PC2, contributing 50% of the total data variability (PC1: 33% and PC2: 17%). The RGB images' score plot shows the five cultivated rice seed varieties clustered to the centre of the axes and overlapping, indicating the cultivated rice seed group has high similarities. Meanwhile, the scores of weedy rice seeds clustered together and fell into the negative quadrant of the PC1 and PC2 axes. However, there was an overlapped area between the weedy rice and cultivated rice seeds, indicating that weedy rice seeds have high similarities to the cultivated rice group.

The loadings plot describes the features' relationships on principal component axes. Based on the factor loadings, PC1 was characterised strongly by morphological features such as the seed maximum Feret diameter (0.148), convex hull perimeter (0.138), major axis length (0.132) and seed perimeter (0.129) on the positive side of the axes, whereas the textural features such as shade R (-0.210), mean B (-0.214), variance R (-0.215) variance G (-0.216) and mean G (-0.218) represented PC1 on the negative side. In contrast, the PC2 was mainly represented by the colour features such as the Mean V (0.272), Mean R (0.272) and Mean G (0.262) on the positive side, while the Std Dev G (-0.273), Std Dev R (-0.277) and Std Dev

V (-0.278) had strong negative loading on PC2. Features close to the axes' centre have no or lesser extent in the principal components.

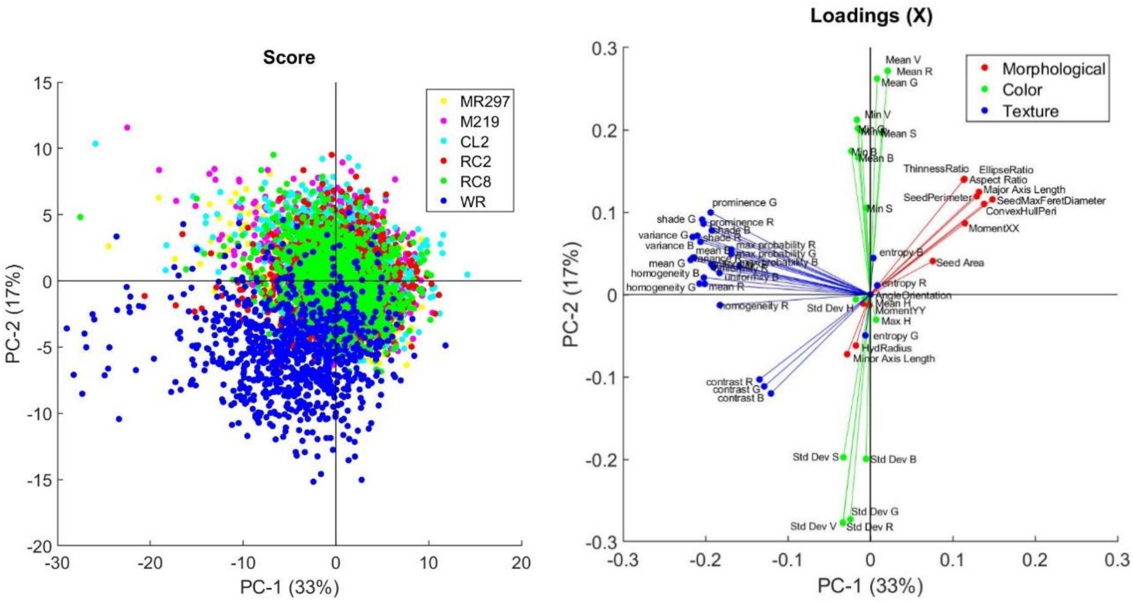

Figure 2. Principal Component Analysis of the RGB image dataset (**a**) Score plot; (**b**) Loadings plot

From Figure 2, it is suggested that the weedy rice seed had a strong relationship to the colour features and textural features on the negative quadrant of PC1 and PC2. Weedy rice seed's characteristics showed no correlation to the morphological and texture features in the positive quadrant of PC1 and PC2. Meanwhile, the cultivated rice seedwas mainly characterised by the morphological mean value of the colour and most of the textural features of the RGB image, as shown in the positive quadrant of PC1 and PC2.

4. Discussion

The PCA function is to explore the dataset distributions using the score plot through the sample patterns on the principal components and it shows the sample differences or similarities (Bro & Smilde, 2014). The closer the samples on the score plot, the more similar they are, as seen in Figure 2a. High similarities of the seed features, as shown by the overlapped area in the score plot, exhibited in the weedy rice seed, are understood as the hybridisation of the wild Oryza population and the subspecies of the Indica rice (Sudianto *et al.*, 2016; Song *et al.*, 2014; Ruzmi *et al.*, 2021). The overlapping among the two seed groups results in complex discriminative rules by the classifier (Sun *et al.*, 2009). Further classification of the weedy rice cultivated by Ruslan *et al.* (2022) resulted in lower sensitivity in weedy rice classification, in comparison to the cultivated rice seed groups. It can be deduced that the similarity of the weedy rice and cultivated rice seeds in this study was high, resulting in a higher possibility of misclassification of weedy rice seeds due to the vague distinction between cultivated rice and weedy rice seeds. It also has been proved by Koklu & Ozkan (2020) that obvious distinction between dry bean sample varieties, especially in shape and size, would lead to a high success rate (>90%) in binary classification.

5. Conclusions

This study has presented the characteristics of weedy rice seed using PCA visualisation and observation on the score and loading plot using three major parameters, as well as the morphology, colour, and texture of RGB image. The PCA has concluded that the overlapped area between weedy rice and cultivated rice seeds was mainly due to the hybridisation of the wild Oryza and the Indica rice subspecies.

Author Contributions: Conceptualisation, methodology, software, analysis, investigation, data curation, writing original draft: Rashidah Ruslan; Supervision, resources, writing—review and editing, Siti Kharunniza-Bejo and Mahirah Jahari; Supervision; Ibni Hajar Rukunuddin.

Funding: This work was funded by the Ministry of Higher Education Malaysia under the Prototype Research Grants Scheme (PRGS) with grant number 9013-00032.

Acknowledgements: The authors would like to thank the Department of Agriculture, Teluk Cengai, Kedah for the resources and seed samples.

Conflicts of Interest: The authors declare no conflict of interest.

References

Abdullah, M. Z., Guan, L. C., Mohd Azemi, B. M. N. (2001). Stepwise Discriminant Analysis for colour grading of oil palm using machine vision system. *Food and Bioproducts Processing*, *79*, 223–231.

Ahmed, Q. N., Hussain Zainudin, P. M. D., & Othman, A. S. (2012). Comparative study on vegetative and reproductive development between weedy rice morphotypes and commercial rice varieties in Perak, Malaysia. *Tropical Life Sciences Research*, *23*, 17–25.

Ansari, N., Ratri, S. S., Jahan, A., *et al*. (2021). Inspection of paddy seed varietal purity using machine vision and multivariate analysis. *Journal of Agriculture and Food Research*, *3*, 100109.

Aznan, A., Viejo G. C., Pang, A., *et al*. (2021). Computer vision and machine learning analysis of commercial rice grains : A potential digital approach for consumer perception studies. *Sensors*, *21*, 6354.

Bro, R., & Smilde, A. K. (2014). Principal Component Analysis. *Analytical Methods*, *6*, 2812.

Huang, K. Y., & Chien, M. C. (2017). A novel method of identifying paddy seed varieties. *Sensors (Switzerland)*, *17*.

Karim, R. S. M., Man, A. B., & Sahid, I. B. (2004). Weed problems and their management in rice fields of Malaysia: An overview. *Weed Biology and Management*.*4*, 177–186

Koklu, M., & Ozkan, I. A. (2020). Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, *174*, 105507.

Kuo, T. Y., Chung, C. L., Chen, S. Y., *et al*. (2016). Identifying rice grains using image analysis and sparserepresentation-based classification. *Computers and Electronics in Agriculture*, *127*, 716–725.

Maione, C., & Barbosa, R. M. (2019). Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. *Critical Reviews in Food Science and Nutrition*, *59*, 1868–1879.

Nadir, S., Xiong, H. B., Zhu, Q., *et al*. (2017). Weedy rice in sustainable rice production. A review. *Agronomy for Sustainable Development*, *37*.

Qiu, Z., Chen, J., Zhao, Y., *et al*. (2018). Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Applied Sciences (Switzerland)*, *8*, 1–12.

Ruslan, R., Khairunniza-Bejo, S., Jahari, M., *et al*. (2022). Weedy rice classification using image processing and a machine learning approach. *Agriculture (Switzerland)*, *12*, 1–15.

Ruzmi, R., Ahmad-Hamdani, M. S., Abidin, M. Z. Z., *et al*. (2021). Evolution of imidazolinone-resistant weedy rice in Malaysia: The current status. *Weed Science*, *69*, 598–608.

Song, B. K., Chuah, T. S., Tam, S. M., *et al*. (2014). Malaysian weedy rice shows its true stripes: Wild Oryza and elite rice cultivars shape agricultural weed evolution in Southeast Asia. *Molecular Ecology*, *23*, 5003–5017.

Sudianto, E., Neik, T.X., Tam, S. M., *et al*. (2016). Morphology of Malaysian weedy rice (Oryza sativa): Diversity, origin and implications for weed management . *Weed Science*, *64*, 501–512.

Sun, J., Qian, Q., Ma, D. R., *et al*. (2013). Introgression and selection shaping the genome and adaptive loci of weedy rice in Northern China. *New Phytologist*, *197*, 290–299.

Sun, Y., Wong, A. K. C., & Kamel, M. S. (2009). Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, *23*, 687–719.

Zhao, C., Xu, W., Song, X., *et al*. (2018). Early flowering and rapid grain filling determine early maturity and escape from harvesting in weedy rice. *Pest Management Science*, *74*, 465–476.