Selection of Morphological Features in Classifying Weedy Rice and Rice Seed Varieties using Discriminant Function Analysis

R Ruslan¹,², S K Bejo¹, I H Rukunuddin² and A A Aznan²

¹Department of Biological and Agricultural Engineering, Universiti Putra Malaysia, UPM 43400 Serdang, Selangor Malaysia
²School of Bioprocess Engineering, Universiti Malaysia Perlis, UniMAP 02600 Arau, Perlis Malaysia
³Center of Excellence for Advanced Sensor Technology (CEASTech) Universiti Malaysia Perlis, UniMAP 02600 Arau, Perlis Malaysia

Abstract. A comparison on classification accuracy and morphological features selection between stepwise discriminant function analysis (DFA) and enter all variables DFA were made for classification of Malaysian rice seed varieties namely MR297, MR263, MR284, MR219 and a group of weedy rice. Eighteen morphological features were used for the discriminant analysis. The classification accuracies of MR297 and MR263 against weedy rice group were maintained at 99.1% and 98.9% respectively, either using stepwise DFA or normal DFA. The classification accuracy for MR284 decrease from 95.0% to 93.7% using stepwise DFA with the reduction to five morphological features selected in the analysis. However for MR219, the classification accuracy increased by 0.3% using optimized features in stepwise DFA. Thus it can be concluded the optimization of morphological features through stepwise DFA classification does not necessarily increase the accuracy for discriminating weedy rice and Malaysian rice seed varieties.

1. Introduction
Production of rice seed in Malaysia is greatly depends on the purity of the cultivated rice seed produced under the certified rice seed program monitored by the Department of Agriculture (DOA). Under this program, the farmers are ensured to receive high quality seeds by monitoring the seed production activities starting from planting to harvesting, processing and seed sampling in the laboratory [1]. Rice seed sampling is one of the crucial part where the processed rice seeds are tested for germination rate, moisture content, and finding foreign/unwanted seed. Normally, seed producers having difficulty in getting clearance to meet the specification from DOA where they fail the test in finding foreign/unwanted seed. Thus, the purpose of this project is to replace manual process of distinguishing the unwanted seed or known as weedy rice from the cultivated rice seed samples. With the advancement of technology, a machine vision-based inspection system was developed for this project to identify the cultivated rice seeds and weedy rice seeds.
2. Related Research

Substantial work related to machine vision-based in agriculture especially in the rice seed industry has been carried out. Common features used in classifying the rice grains are morphological [2][3], color, textural and combination of the said features [4]. Some researchers such as [5] identify the sterile lemma traits in discriminating between rice seed cultivars. The classifiers tested for image classification were done using either parametric or non-parametric such as support vector machine (SVM), neural network, k-nearest neighbour and discriminant function analysis (DFA). The classification accuracy up to 85% were achieved in classifying rice seed cultivars. Other works involving rice seed classification were to discriminate between weedy rice and rice seed cultivars [6]. [4] extracted texture and shape feature of four Indian rice variety using vision based and image processing and evaluated the classification accuracy using neural network. The classification accuracy were highest for shape features, followed by texture-n-shape and texture features at 88%, 87.27% and 82.61% respectively. Work by [5] used microscopic imaging to identify 45 rice seed cultivar using morphological, colors and sterile lemma traits of the rice grains with overall classification accuracy reached to 85.02% by support vector machine (SVM) classifier.

3. Materials and Method

3.1. Machine Vision Setup

A machine vision-based setup consist of a CMOS 6MP 1/1.8” GigE camera (MVCA060-10GC, HIK Vision) and 6MP 25mm focal length lens (MVL-HF2528M-6MP, HIK Vision) was setup for this project. The area scan camera resolution is 3072 x 2048 and has sensor size of 2.4 µm x 2.4 µm per pixel. A laptop with 12Gb RAM memory and 2.8 GHz (Intel i7) processing speed was attached to the camera through Ethernet cable CAT6 with speeds up to 1000 Mb/s. Image processing and analysis was programmed in LabVIEW 2016 (National Instruments, Texas USA) visual programming language. A low angled LED lighting (TMS Lite) was used and provides even and controlled lighting distribution on the seed samples. The seed sample was kept at a distance of 13.7 cm from the lens and camera.

3.2. Rice Seed Samples

Cultivated rice seed samples used in this project were supplied from Rice Seed Bank (MARDI Seberang Perai, Malaysia). Four rice seed varieties which in demand every planting seasons (MR297, MR263, MR284 and MR219) and two weedy rice variants (the unwanted/off type rice seed similar to cultivated rice) gathered from plantation area were used in this project. Figure 1 displays sample images of the cultivated rice seed and weedy rice variants. The shape of the rice seed varieties is almost similar to each other and weedy rice variants used in this project is slightly smaller than cultivated rice seed varieties and have an awn.

![Rice seed varieties](a) MR219 (b) MR284 (c) Weedy Rice Variant A (d) MR297 (e) MR263 (f) Weedy Rice Variant B

**Figure 1.** Rice seed variety and weedy rice variants (a) MR219 (b)MR284 (c) Weedy Rice Variant A (d) MR297 (e) MR263 (f) Weedy Rice Variant B
3.3. Morphological features
Feature extraction was programmed in LabVIEW environment and IMAQ Particle Analysis VI was used. The following morphological features [7] as depicted in table 1 were extracted from labelled images of individual seed kernels:

| No | Features                     | Unit   | Descriptions                                                      |
|----|------------------------------|--------|-------------------------------------------------------------------|
| 1  | Length                       | mm     | Length of the bounding rectangular width of the seed kernel       |
| 2  | Width                        | mm     | Length of the bounding rectangular height of the seed kernel      |
| 3  | Area                         | mm     | Number of pixels in the area of the seed kernel                   |
| 4  | Perimeter                    | mm     | Length of the pixel boundary of the seed kernel                   |
| 5  | Convex Hull Perimeter        | mm     | Length of the smallest convex polygon containing all pixels in the seed kernel |
| 6  | Seed Maximum Feret Diameter  | mm     | The longest length of the line segment connecting the two end points on the seed kernel |
| 7  | Major Axis Length            | mm     | Length of the major axis of the ellipse that has the same area and same perimeter as the kernel |
| 8  | Minor Axis Length            | mm     | Length of the minor axis of the ellipse that has the same area and same perimeter as the kernel |
| 9  | Aspect Ratio                 |        | - Major axis length/ Minor axis length                            |
| 10 | Rectangular Aspect Ratio     |        | - Length/Width                                                   |
| 11 | Thinness Ratio               |        | - (Seed Perimeter)/(Seed Area*π)                                 |
| 12 | Area ratio                   |        | - (Length/Width)* Seed Area                                      |
| 13 | Ellipse Ratio                |        | - Major Axis Length/Minor Axis Length                             |
| 14 | Rectangular Diagonal         | mm     | Length of the opposite corner of the bounding rectangular of the seed kernel |
| 15 | Hydraulic Radius             | mm     | Disk area/disk perimeter where the disk radius is the distance between a pixel boundary to the centroid of the seed kernel |
| 16 | Angle Orientation            | radian | The angle of the line that passes through the seed centroid about which the particle has the lowest moment of inertia. |
| 17 | Moment XX                    |        | - Moment on the X direction to the particle center of mass       |
| 18 | Moment YY                    |        | - Moment on the Y direction to the particle center of mass       |

3.4. Image Processing and Analysis
Simple image processing technique were employed to analysed the seed kernel images and features extraction. Figure 2 shows the flowchart for the image processing procedure run in the LabVIEW program.

Image analyses were performed based on non-parametric analysis. Discriminant analysis was conducted using methods (i) enter all independent variables into Discriminant Function Analysis (DFA) and (ii) stepwise DFA (SPSS). The analyses were arranged one-to-one scheme between weedy rice group and each cultivated rice seed. For both cross validation analysis, features selection were based on the minimization of the Wilks’ Lambda. Morphological features that minimize the Wilks’ Lambda value is selected as the predictors for the function equation. Criteria for entry and removal of the features into the predictors equation (F criteria) was set at 0.05 and 0.10, respectively.
4. Result and discussion

4.1. Morphological features selection using DFA and stepwise DFA

Stepwise DFA technique removes or retained independent features based on the reduction of overall Wilks’ Lambda in each entry while the normal DFA would enter all variables at once. In general, overall number of features selected for the discriminant function decreased from 12 to six, seven, five and eight for MR297, MR263, MR284 and MR219, respectively as indicated in table 2. In normal DFA technique, all 12 features selected for the function generation were the same in the four seed varieties. Meanwhile, the most significant features in stepwise DFA were convex hull perimeter, minor axis length and area ratio as it was selected by three rice seed varieties out of four for classification function generation.

Table 2. Morphological features selection using stepwise and normal discriminant function analysis.

| No | Morphological Features | Rice Seed Varieties | Number of times feature appeared in stepwise DFA |
|----|------------------------|---------------------|-----------------------------------------------|
| 1. | Width                  | MR297 DFA | MR297 Stepwise DFA | MR263 DFA | MR263 Stepwise DFA | MR284 DFA | MR284 Stepwise DFA | MR219 DFA | MR219 Stepwise DFA |
| 2. | Length                 | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 3. | Seed Area              | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 4. | Convex Hull Perimeter  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 5. | Seed Perimeter         | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 6. | Seed Maximum Feret Diameter | ✔  | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 7. | Major Axis Length      | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 8. | Minor Axis Length      | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 9. | Aspect ratio           | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 10.| Rectangular Diagonal   | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 11.| Thinness Ratio         | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 12.| Hydraulic Radius       | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 13.| Angle Orientation      | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 14.| Rectangular Aspect Ratio| ✔      | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 15.| Moment XX              | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 16.| Moment YY              | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 17.| Ellipse Ratio          | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |
| 18.| Area Ratio             | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  | ✔        | ✔                  |

MR297 rice seed group has shown highest discriminatory ability among other seed groups by having smallest values of Wilks Lambda. High variance in the independent features selected was explained by the function indicated from the large eigenvalues. The lowest classification accuracy among the rice seed varieties groups was shown by MR284 due to lower number of features (five out of 18) selected in the stepwise DFA, thus increased the Wilk’s Lambda value to 0.303.

The classification accuracies of MR297 and MR263 with weedy rice group were maintained at 99.1% and 98.9% respectively, either using stepwise DFA or not. The classification accuracy for MR284
decreased from 95.0% to 93.7% using stepwise DFA with the reduction to five morphological features selected in the function. In this case, the classifier performance decline when lower number of features were selected by stepwise DFA. However for MR219, the classification accuracy increased by 0.3% using optimized features in stepwise DFA.

5. Conclusion
It can be deduced from the results, either using optimized features or not, the generation of the function coefficient greatly depending on the discriminatory ability from Wilks’ Lambda reduction during entry of the independent features. The most significant features selected in stepwise DFA were convex hull perimeter, minor axis length and area ratio.

6. References

[1] Ismail A T and Said A R 2012 Certified Paddy Seed Production and Processing, *Int. Conf. on Agricultural and Food Engineering for Life* (Cafei2012) pp1-5

[2] Gayathridevi T, Neelamegam P and Sudha S Machine Vision based Quality Analysis of Rice Grains

[3] Guzman J D and Peralta E K 2008 Classification of Philippine Rice Grains Using Machine Vision and Artificial Neural Networks *Proc. World Conf. on Agricultural and IT* (Nagatsuka T:Tokyo) Tokyo University of Agriculture pp 41-48

[4] Chaugule A and Mali S M 2014 Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties Journal of Engineering Hindawi Publishing Corporation Volume pp 8

[5] T Y Kuo, S Y Chen, H A Lin, C L Chung, Y F Kuo 2015 Identifying Rice Grains of Various Cultivars Using Machine Vision ASABE Paper No. 152188279 St. Joseph, Mich.: ASABE

[6] A.A. Aznan, I.H. Rukunudin, A.Y.M. Shakaff, R. Ruslan, A. Zakaria and F.S.A. Saad 2014 Application of Image Processing Technique to Extract Morphological Characteristics of Weedy Rice Seeds Variants for Malaysian Seed Industry”. *Adv. Environ. Biol.*, vol 8(22), pp. 112-115

[7] IMAQ Vision for LabVIEW User Manual 2000 National Instruments Corp, Austin, Texas