J. Trop. Agric. and Fd. Sc. 37(2)(2009): 241–248

Immature paddy quantity determination using image processing and analysis techniques

(Penentuan kuantiti padi tidak matang dengan menggunakan teknik pemprosesan dan penganalisisan imej)

C.C. Teoh* and B.H. Abu Bakar*

Keywords: immature paddy detection, paddy weight prediction, image processing and analysis

Abstract

Image processing and analysis techniques that enable determination of percentage immature paddy in a sample were developed. The red, green and blue (RGB) images captured from digital colour camera were transformed to the principle component analysis (PCA) images for detecting immature paddy. The transformed images successfully identified the immature paddy with mean accuracy of 99.8%. An immature paddy weight prediction model was developed for calculating the percentage immature paddy. Result showed that the model is capable to predict the percentage of immature paddy with mean accuracy of 94.45%.

Introduction

Rice constitutes the world's principle source of food, being the basic grain for the planet's largest population. For tropical Asians, it is the staple food and is the major source of dietary energy and protein. In Southeast Asia alone, rice is the staple food for 80% of the population (Yadav and Jindal 2001). Thus, the selection of good quality paddy is essential to produce high quantity and quality milled rice.

Immature paddy is one of the important parameters to be considered in the milling processes because it is chalky and produces broken rice. According to Rohani et al. (1992), the amount of immature paddy in the harvested paddy should be 5–10%. At this level of percentage, the grain will not crack easily and will give good milling yield.

Presently, the percentage of immature paddy in a sample is determined by human inspection, separation, weighing and calculation. These manual operations are tedious, time consuming, labour intensive and costly.

Over the past decade, advances in hardware and software for digital image processing and analysis have motivated several studies on the development of vision system to evaluate the quality of diverse and processed foods (Gerrard et al. 1996; Locht et al. 1997).

Computer vision has been used in grain quality inspection for many years. An early study by Zayas et al. (1989), machine vision was used to identify different varieties of wheat and to discriminate wheat from non-wheat components.

*Mechanization and Automation Research Centre, MARDI Headquarters, Serdang, P.O. Box 12301, 50774 Kuala Lumpur, Malaysia
Authors' full names: Teoh Chin Chuang and Badril Hisham Abu Bakar
E-mail: cchin@mardi.gov.my

The relationship between colour and texture features of wheat samples to scab-infection rate was studied using machine vision and neural network methods by Ruan et al. (1997). It was found that the infection rates estimated by the system followed the actual ones with a correlation coefficient of 0.97 with human panel assessment and maximum and mean absolute errors of 5% and 2% respectively.

Measurements of kernel length, width and projected area independent of kernel orientation were performed using machine vision by Ni, Paulsen and Reid (1997). The algorithm accuracy was between 0.86 and 0.89 measured by the correlation coefficient between predicted results and actual sieving for a 500 g sample.

Steenhoek and Precetti (2000) performed a study to evaluate the concept of two-dimensional image analysis for classification of maize kernels according to size category. A total of 320 maize kernels were categorised into one of 16 size categories based on degree of roundness and flatness. Classification accuracy of both machine vision and screen systems was above 96% for round-hole analysis. However, sizing accuracy for flatness was less than 80%.

The automatic inspection of 600 corn kernels was also performed by Ni, Paulsen, Liao et al. (1997) using machine vision. For kernel identification, on-line tests had successful classification rates of 91% and 94% for whole and broken kernels respectively.

Liu and Paulsen (1997) developed a digital image analysis method for measuring the degree of milling rice. They compared the method with conventional chemical analysis and obtained a coefficient correlation of 0.9819 for the 680 samples tested.

Wan et al. (2000) employed three on-line classification methods for rice quality inspection (i.e. range selection, neural network and hybrid algorithms). The highest recorded on-line classification accuracy was around 91% at a rate of over 1,200 kernels/min. The range selection method achieved this accuracy but required time-consuming and complicated adjustment.

The objective of this paper was to introduce image processing and analysis techniques which could accurately determine the percentage of immature paddy of a sample.

Materials and methods

Paddy samples

A total of 40 and 60 paddy samples (MR 220) with green and yellow colours were obtained from Sg. Burong, Tanjung Karang, Selangor on 21 June 2007 and FELCRA, Seberang Perak, Perak on 25 December 2007 respectively. The immature (green) and mature (yellow) paddy of the samples were separated and weighed.

Image acquisition

An image acquisition system (*Plate 1*) was designed to allow each paddy of a sample to be placed individually under the lighting and camera for imaging. A Linux Model colour camera was used to grab red, green and blue (RGB) images of paddy. The distance between object and the camera was 30 cm. The whole paddy samples obtained from the two study sites were run through the system to capture the images. The resolution of each captured image was 70 pixel/mm^2. An image processing PCI software was used to process and analysis all paddy sample images.

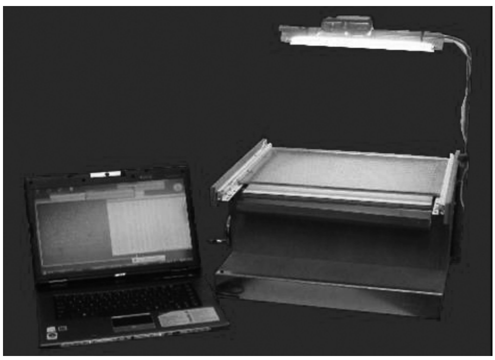


Plate 1. Image acquisition system

Immature paddy detection

Immature paddy detection was performed by transforming the RGB images to principle component analysis (PCA) images.

The PCA is a statistical technique that transforms a multivariate data set consisting of intercorrelated variables into a data set consisting of variables that are uncorrelated linear combinations of the original variables. The transformed variables are referred to as principle components (PCs). Mathematically, PCA is a matrix diagonalization process that determines the mixture eigenvectors, with the eigen values corresponding to the relative presence of each of these mixtures in the image. Generally, the first PC expresses the maximum possible proportion of the variance in the original data set; subsequent PCs account for successively smaller proportions of the remaining variance (Ingebritsen and Lyon 1985).

In this study, the transformed PC images were analysed to select the best image for detecting immature paddy.

Percentage of immature paddy determination

An immature paddy weight prediction model was developed for calculating the percentage of immature paddy using the formula:

$$\frac{\text{Predicted weight of immature paddy}}{\text{Total weight of paddy sample}} \times 100\%$$

The model is based on the outline area of immature paddy to predict paddy weight.

A simple thresholding algorithm was used to segment and outline the paddy object of a paddy sample image. The thresholding algorithm is shown below:

$$g(x,y) = \begin{cases} 0, & \text{if } f(x,y) \geq T_2 \\ 1, & \text{otherwise} \end{cases}$$

where $g(x,y)$ is the image after thresholding, $f(x,y)$ is the original paddy image and T_2 is the threshold. All pixels of image with grey level values equal or greater than T_2 were labelled as background pixels (0), while pixels with grey level values smaller than T_2 were classified as paddy pixels (1).

To outline the paddy object, the thresholded image was further processed by scanning the image until the edges of all paddy objects were outlined. The area of each outline paddy object was measured by the software for developing an immature paddy weight prediction model.

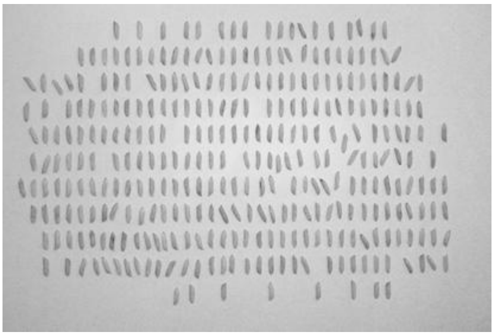
A method of linear regression analysis of SAS 8.0 software was used to analyse the relationship between area outline and weight of each immature paddy sample. If the linear regression has a high correlation coefficient value between the two variables, then a linear regression model can predict accurately the weight of immature paddy (dependent variable) by the measurement of its area outline (independent variable) which is fitted to the model.

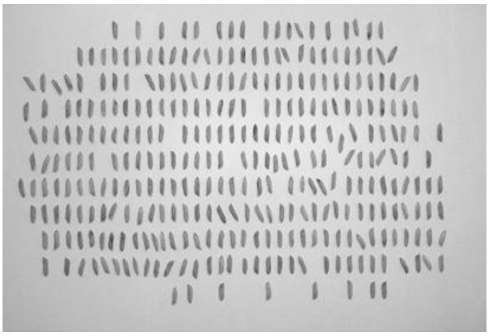
A 95% confidence and prediction intervals graph was used to prove the linear model adequately fits the data and a plot of residuals versus independent values was used to detect the violation of assumptions of the linear model is appropriate. Generally, when the linear model is assumed as appropriate, the residuals must be independent and normally distributed with the same variance everywhere.

Results and discussion

To detect immature paddy, the R, G and B images (*Plate 2*) were transformed to PC1, PC2 and PC3 images (*Plate 3*). The three transformed images were analysed by visualisation to select the best image to detect the immature paddy. It can be seen that the PC3 image was capable to differentiate the immature paddy (bright object) from the mature paddy (dark object) compared to PC1 and PC2 images.

An analysis of grey level value of immature and mature paddy for R, G, B and PC3 images were studied. The statistical analysis results (*Figure 1*) indicated that the R, B and PC3 images were significantly different (p <0.05) except G image and the highest different mean grey level was found


R image


G image

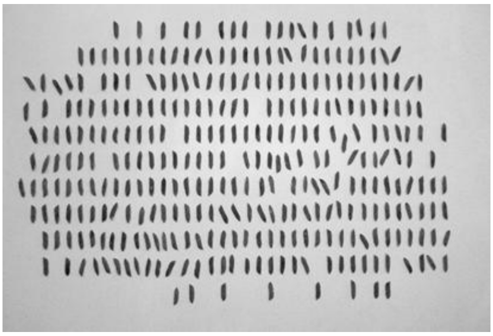
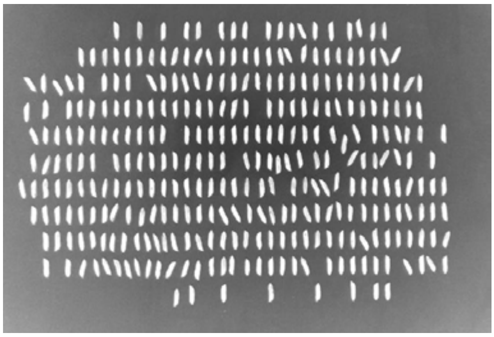
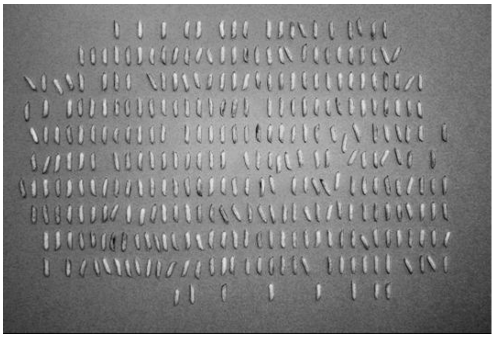

B image

Plate 2. R, G and B images


PC1 image


PC2 image

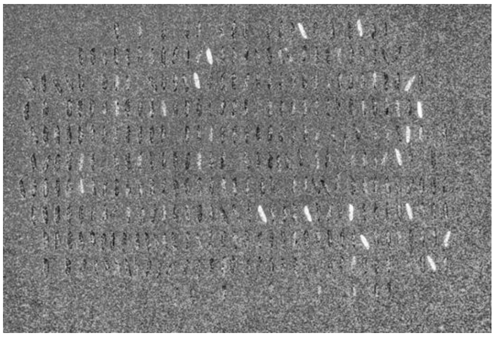

PC3 image

Plate 3. PC1, PC2 and PC3 images

in the PC3 image. Therefore, the PC3 image was selected for detecting immature paddy.

Table 1 shows the accuracy of immature paddy detection of 10 PC3 images that were transformed from the R, G and B images obtained from FELCRA. The accuracy ranged from 98% to 100% with mean value of 99.8%. The result showed that the transformed PC3 image is capable to detect immature paddy with high accuracy.

A total of 40 immature paddy samples obtained from Sg. Burong were used to develop a linear regression model for predicting immature paddy weight based on paddy outline area. The predicted weight was used to calculate the percentage of immature paddy in a sample. *Plate 4* presents one of the immature paddy images captured by the image acquisition system. The paddy objects of the image were segmented from the background by the thresholding algorithm (*Plate 5*). The edges of all paddy objects in the thresholded image were outlined by the scanning

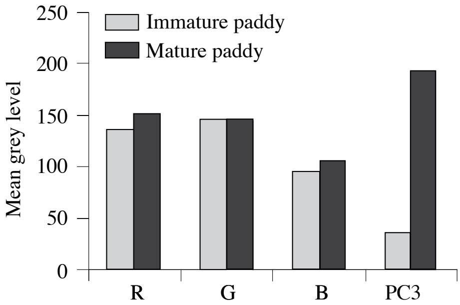


Figure 1. Comparison of mean grey level of immature paddy with mature paddy for R, G, B and PC3 images

Table 1. Immature paddy detection accuracy of 10 image samples

Images	Actual number of immature paddy	Identified number of immature paddy	Accuracy (%)
1	11	11	100
2	18	18	100
3	26	26	100
4	35	35	100
5	40	40	100
6	54	53	98
7	52	52	100
8	65	65	100
9	67	67	100
10	77	77	100

process (*Plate 6*). Areas of the all paddy outlines were measured by the software for developing an immature paddy weight prediction model.

Table 2 shows the immature paddy weight and outline areas of 20 sub-group samples. The outline areas ranged 392–7,689 mm^2, while the paddy weights ranged 0.5–10 g. A simple analysis indicated that the weight and outline area has a strength of relationship, when the weight increases the outline area also increases. Mean (μ) and standard deviation (s) for the weights were 5.3 and 3.0 respectively, while for the outline areas were 4,123.56 and 2,276.91 respectively. The calculated of 95% confidence interval for μ of the weight and the areas ranged 3.87–6.63 and 3,057.93–5,189.19 respectively.

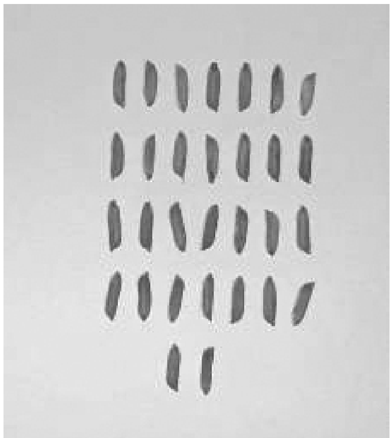


Plate 4. Immature paddy sample image

Plate 5. Segmented image

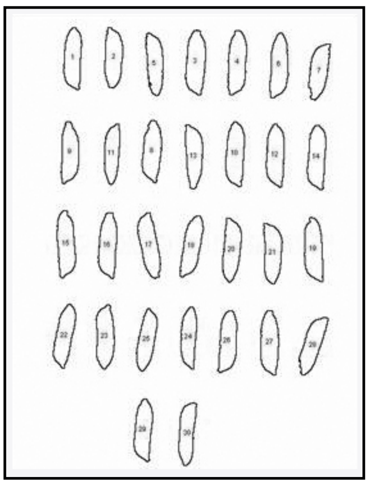


Plate 6. Paddy outlines image

Table 2. Immature paddy weight and average outline areas of 20 sub-group weights paddy samples

Paddy weight (g)	Paddy outline area (mm^2)		
	Set 1	Set 2	Average
0.5	364.17	420.45	392.31
1.0	776.06	819.78	797.92
1.5	1224.16	1226.86	1225.51
2.0	1652.53	1590.93	1621.73
2.5	2057.26	1963.57	2010.41
3.0	2485.39	2366.43	2425.91
3.5	2805.88	2686.69	2746.29
4.0	3257.55	3081.30	3169.43
4.5	3721.79	3395.10	3558.44
5.0	4194.25	3883.91	4039.08
5.5	4452.57	4167.61	4310.09
6.0	4908.94	4659.98	4784.46
6.5	5307.13	4992.22	5149.68
7.0	5879.60	5291.79	5585.70
7.5	5950.54	5632.88	5791.71
8.0	6439.72	5976.07	6207.89
8.5	6823.76	6432.25	6628.01
9.0	7265.60	6727.56	6996.58
9.5	7517.05	7163.28	7340.16
10.0	7800.66	7579.07	7689.87

The weights were correlated with the outline areas using linear regression analysis method. The best fit line that described the relationship between the two variables is shown in *Figure 2*. The scatterplot indicated a good dependency with high liner coefficient correlation (r) value of 0.9995 of the two variables. As a result, a linear regression model for predicting the weight of immature paddy using the measured paddy area outline can be developed by fitting a line to the data. The developed model is given below:

Predicted weight = 0.0013* paddy outline area – 0.1054

The plot also presented that a good 95% confidence and prediction intervals result for prediction of immature paddy weight. The prediction bands are wider than the corresponding confidence bands to allow for the fact that the linear model predicting the value of a random variable rather than

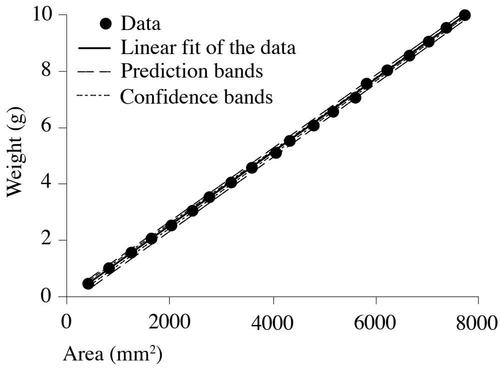


Figure 2. The dependency between area and weight

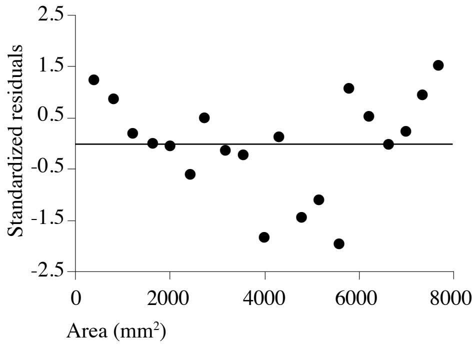


Figure 3. The residuals (error) distribution

estimating a parameter. The 95% confident interval for intercept and slope of linear model were –0.1818 to –0.0289 and 0.00128 to 0.00132 respectively.

Figure 3 illustrates a plot of residuals versus independent values. It can be seen that the plot shows homogeneous error variances and does not show any pattern. This gave confidence regarding the normal distribution of the data. The result from the plot also showed that there are no extremely large residuals and that there is no trend in the residuals to indicate that the linear model is inappropriate.

To verify the developed linear model for determining the percentage of immature paddy in a paddy sample, 60 R, G and B paddy images obtained from FELCRA were used. *Table 3* summarises the actual percentage, predicted percentage and accuracy percentage prediction of immature

Table 3. Accuracy of predicted percentage of immature paddy

Samples	% of immature paddy		Accuracy %	Samples	% of immature paddy		Accuracy %
	Actual	Predicted			Actual	Predicted	
1	2.0	1.9	95.0	31	13.6	13.9	97.9
2	4.0	4.3	92.5	32	14.0	13.2	94.3
3	5.0	5.8	84.0	33	14.4	14.0	97.5
4	6.0	4.8	80.0	34	14.4	14.9	96.7
5	6.0	6.5	91.7	35	15.2	15.8	96.3
6	8.0	7.8	97.5	36	15.6	15.4	99.0
7	9.0	8.9	98.9	37	16.0	15.9	99.5
8	10.0	9.6	96.0	38	16.0	17.0	93.8
9	12.0	11.3	94.2	39	16.8	17.0	98.8
10	12.0	11.8	98.3	40	16.8	17.4	96.7
11	8.0	8.33	95.8	41	14.7	14.8	99.1
12	9.3	9.73	95.7	42	15.0	15.2	98.9
13	10.0	9.6	96.0	43	15.0	16.1	92.9
14	10.7	11.3	93.8	44	15.3	15.7	97.4
15	12.0	11.5	96.1	45	16.0	16.5	96.7
16	12.0	11.5	95.6	46	16.0	16.8	95.2
17	12.0	12.7	93.9	47	16.7	17.6	94.4
18	13.3	14.6	90.5	48	17.0	18.0	94.3
19	14.7	16.1	90.0	49	17.7	17.6	99.8
20	16.0	15.5	96.7	50	18.7	18.8	99.3
21	12.0	13.1	90.8	51	17.4	17.3	99.3
22	12.0	13.2	90.4	52	19.4	17.3	89.3
23	13.0	14.7	87.3	53	21.4	20.6	96.0
24	13.5	13.3	98.5	54	24.3	22.6	93.2
25	14.0	16.1	85.4	55	25.7	24.2	94.2
26	14.5	14.3	98.3	56	27.4	25.4	92.5
27	15.0	13.9	92.3	57	29.4	26.0	88.3
28	15.0	17.0	87.0	58	30.9	27.7	89.7
29	16.0	18.2	86.6	59	32.9	30.3	92.2
30	16.5	15.5	93.9	60	34.9	35.0	99.6

paddy. The percentage immature paddy prediction accuracy ranged from 80% to 99.8% with mean value of 94.3%. The result indicated that the percentage of immature paddy was predicted with a high degree of accuracy. Therefore, the developed linear model is capable to be used to predict the immature paddy weight for determining percentage of immature paddy in a sample.

Conclusion

An image processing and analysis techniques were developed for detecting immature paddy and determining percentage of immature paddy in a paddy sample. Based on the transformation of R, G and B images to principle component analysis (PCA) images, the transformed PC3 image provided a successful detection of immature paddy with mean accuracy of 99.8%. The percentage of immature paddy can be predicted by the developed weight prediction model up to 94.45% mean accuracy. The two accuracy results indicated that, the image processing and analysis techniques demonstrated potential for replacing the manual procedure (i.e. human inspection, separation, weighing and calculation) to determine percentage of immature paddy in a sample.

References

Gerrard, D.E., Goa, X. and Tan, J. (1996). Beef marbling and colour score determination by image processing. *Journal of Food Science* 61: 145–148

Ingebritsen, S.E. and Lyon, R.J.P. (1985). Principle components analysis of multitemporal image pairs. *International Journal of Remote Sensing* 6: 687–696

Liu, J. and Paulsen, M.R. (1997). Corn whiteness measurement and classification using machine vision. In: *1997 ASAE Annual International Meeting Technical Papers,* (Paper No. 973045). Michigan: ASAE

Locht, P., Thomsen, K. and Mikkelsen, P. (1997). *Full color image analysis as a tool for quality control and process development in the food industry.* (Paper No. 973006). Michigan: ASAE

Ni, B., Paulsen, M.R., Liao, K. and Reid, J.F. (1997). Design of an automated corn kernel inspection system for machine vision. *Transactions of the ASAE* 40(2): 491–497

Ni, B., Paulsen, M.R. and Reid, J.F. (1997). Size grading of corn kernels with machine vision. In: *1997 ASAE Annual International Meeting Technical Papers,* (Paper No. 973046). Michigan: ASAE

Rohani, M.Y., Samsuddin, A., Ajimilah, N.H., Abdullah, A. and Hashifah, M.A. (1992). *Cadangan spesifikasi dan gred padi untuk industri padi negara*. (MARDI Report No. 158). Serdang: MARDI

Ruan, R., Ning, S., Ning, A., Jones, R. and Chen, P.L. (1997). Estimation of scabby wheat incident rate using machine vision and neural network. In: *1997 ASAE Annual International Meeting Technical Papers*, (Paper No. 973042). Michigan: ASAE

Steenhoek, L. and Precetti, C. (2000). Vision sizing of seed corn. In: *2000 ASAE Annual International Meeting,* (Paper No. 003095). Michigan: ASAE

Wan, Y.N., Lin, C.M. and Chiou, J.F. (2000). Adaptive classification method for an automatic grain quality inspection system using machine vision and neural network. In: *2000 ASAE Annual International Meeting,* (Paper No. 003094). Michigan: ASAE

Yadav, B.K. and Jindal, V.K. (2001). Monitoring milling quality of rice by image analysis. *Computers and Electronics in Agriculture* 33: 19–33

Zayas, I., Pomeranz, Y. and Lai, F.S. (1989). Discrimination of wheat classes and varieties by image analysis. *Cereal Chemistry* 66(3): 233–237

Abstrak

Satu teknik pemprosesan dan penganalisisan imej yang berupaya untuk menentukan peratus padi tidak matang di dalam satu sampel telah dibangunkan. Imej-imej merah, hijau dan biru ditangkap oleh kamera warna berdigit dijelmakan kepada imej-imej *principle component analysis (PCA)*. Imej yang terjelma berjaya mengecam padi tidak matang dengan kejituan purata 99.8%. Satu model ramalan berat padi tidak matang telah dibangunkan untuk mengira peratus padi tidak matang. Keputusan menunjukkan bahawa model tersebut berupaya meramalkan peratus padi tidak matang dengan kejituan purata 94.45%.

Accepted for publication on 19 August 2009