Wavelength Selection for Classifying Paddy Rice by Near-Infrared Spectroscopy

Chang-Chun Liu,[1] Jai-Tsung Shaw,[1,2] Keen-Yik Poong,[1] Mei-Chu Hong,[3] and Ming-Lai Shen[4]

ABSTRACT

Cereal Chem. 83(4):335–339

Using five paddy rice cultivars grown in Central, Eastern, and Southern Taiwan and harvested in the summers of 1997, 1998, and 1999, eight calibrated models were established by discriminant analysis and backpropagation neural network with four wavelength selection methods. Randomly adding 80 samples of the 2000 year crop in the three-crop-year calibrated models for annual recalibration, eight models were used to classify paddy rice harvested in the summer of 2000. With 351 wavelengths of models 1 and 2, the average classification rates by discriminant analysis and backpropagation neural network were 98.1 and 92.5%, respectively. With 69 wavelengths selected by stepwise discrimination of models 3 and 4, the average classification rates by discriminant analysis and backpropagation neural network were 98.5 and 85.5%, respectively. With 69 wavelengths selected by correlation matrix of models 5 and 6, the average classification rates by discriminant analysis and neural network were 72.0 and 72.2%, respectively. With 69 wavelengths from loading values in the first and second principal components of models 7 and 8, the average classification rates by discriminant analysis and neural network were 69.1 and 60.6%, respectively. Model 3 would be recommended for classifying paddy rice to set trading prices because of its highest classification rate (98.5%).

Using neural network and near-infrared transmittance, Song et al (1995) classified single kernels of six cultivars of wheat with an average classification rate of 94.7%. Delwiche et al (1995) classified hard red winter wheat and hard red spring wheat using near-infrared spectroscopy by MLR, PCA, PLSR, and ANN models with classification rates of 95.3, 93.0, 98.0, and 98.3%, respectively. Liu and Shaw (1995) detected rice moisture and protein content by near-infrared spectrophotometry. Li and Shaw (1996) studied data processing affecting NIR calibration curves of major constituents of rough rice taste. Li and Shaw (1997) used near-infrared reflectance spectroscopy of rough rice to develop a fat acidity prediction model for whole-kernel rough rice with a 0.93 validation correlation coefficient and a 0.97 validation correlation coefficient for ground rough rice. Chang et al (2000) used near-infrared spectroscopy to validate six cultivars of paddy rice grown at the Taichung District Agricultural Improvement Station in the summer of 1997 and 1998 using PCA and a backpropagation neural network model with a 89.3% validation rate. Wang et al (2002) classified sound and damaged soybean seeds with 99% classification rate by using near-infrared reflectance spectra. Rittiron et al (2005) used near-infrared transmittance spectroscopy to predict the protein of single milled Japanese rice kernels with 0.96 correlation coefficient. Corbella and Cozzolino (2005) used visible and near-infrared spectroscopy to classify two floral origin honeys by PCA, PLSR, and discriminant analysis with an 85% classification rate. Liu et al (2005) used morphological and color features model to classify five popular cultivars (Tainung Sen 20, Taichung Sen 10, Tainung 67, Taikeng 8, and Taikeng 9) grown in Central, Eastern and Southern Taiwan and harvested in the summers of 1997, 1998, 1999, and 2000 with 90–99% classification rates. The objective of this study was to establish eight different near-infrared spectroscopy models by discriminant analysis and backpropagation neural network with four wavelength selection methods from 1100 to 2500 nm in 3-nm steps and to compare with the morphological and color features model by Liu et al (2005). The best model would be recommended for classifying five paddy rice cultivars to set trading prices.

[1] PhD candidate, professor, and former research assistant, respectively. Dept. of Bio-Industrial Mechatronics Engineering, National Taiwan University.

[2] Corresponding author. Phone and fax: 886-2-33665329. E-mail: m320@ntu.edu.tw

[3] Agronomist, Taichung District Agricultural Improvement Station.

[4] Professor, Dept. of Agronomy, National Taiwan University.

DOI: 10.1094/CC-83-0335

MATERIALS AND METHODS

Five popular paddy rice cultivars (Tainung Sen 20, Taichung Sen 10, Tainung 67, Taikeng 8, and Taikeng 9) were grown in Central, Eastern, and Southern Taiwan and harvested in the summers of 1997, 1998, 1999, and 2000, which were the same samples used in the morphological and color features model by Liu et al (2005). Before the test, these samples were packed in meshed bags, stored at 5°C, and were conditioned to ≈13% moisture content in a controlled chamber maintained at 25°C and 70% rh for four days.

The reflectance (R) was defined as the ratio of energy return from a reference ceramic block. Each absorbance (Log (1/R)) of whole-kernel paddy rice in bulk (8 g) was an average of five measurements by rotating the sample cup at 75°. A Bran+Luebbe InfraAlyzer 500 (Norderstedt, Germany) spectroscopy was used to collect the reflectance spectra of bulk kernel samples from 1100 to 2500 nm in 3-nm steps to give 351 data points per sample.

Removing the outlier samples by Mahalanobis distance >9 divided by the sample size (Vellwman and Welsch 1981), 1,951 samples of four crop years remained from 1997 to 2000. About two-thirds of three crop years (1997-1999) samples or 972 samples were used for training sets, and one-third of three-crop-year (1997-1999) samples or 486 samples were used for validation sets. The 493 samples of the 2000 crop year were input into the three-crop-year calibrated model for classifying. By discriminant analysis (Peng 2000), the average validation and classification rates were 100 and 45.31%, respectively. The low classification rate of the three-crop-year calibrated model could be due to close constituents in five paddy rice cultivars, and annual change of climate, soil, fertilizer, etc.

The starch contents of Tainung Sen 20, Taichung Sen 10, Tainung 67, Taikeng 8, and Taikeng 9 were 20.0, 20.1, 20.0, 20.9, and 19.4%, respectively. The protein contents were 7.4, 6.4, 6.3, 6.2, and 6.6%, respectively. To overcome the annual changes of climate, soil, and fertilizer, etc., the established three-crop-year calibrated model may require annual calibration. The input of 20, 40, and 80 samples of the 2000 crop year selected by random for annual calibration, average classifying rates were 85.0, 95.8, and 98.1%, respectively. A total of 80 samples were ≈8% of three-crop-year sample size in training set. Samples of Tainung Sen 20, Taichung Sen 10, Tainung 67, Taikeng 8, and Taikeng 9 were 210, 210, 208, 214, and 212, respectively, in the training set; 97, 96, 96, 98, and 97 samples, respectively, in the validation set; and 84, 82, 83, 82, and 82 samples, respectively, in the classification set.

From Rencher (1995), discriminant functions of variables were used to describe the differences between two or more groups and classification functions of the variables were used to assign an individual sampling unit to one of the groups for prediction. Using the classification function offered by Matlab 7.0, the discriminant method was used to develop validation and classification programs.

Using sigmoid ($y = 1/(1+\exp(x))$) as a nonlinear transform function, a backpropagation neural network was applied with supervised learning to minimize the global error of the system by modifying weighting in nodes with 50,000 allowable iterations. The Professional II/Plus software (NeuralWare, Pittsburgh, PA), was used to develop validation and classification programs of five paddy rice cultivars.

The 351 variables (absorbance (Log(1/*R*)) were input into a discriminant analysis program to find model 1. The 351 variables were input into a backpropagation neural network for training to find model 2 with 60 nodes in the first hidden layer and 40 nodes in the second hidden layer, 50,000 iterations, an RMS error of 0.11, and a correlation coefficient of 0.78.

To reduce the wavelength number from 351 down to 69 to save computation time and maintaining reasonable classification rate, stepwise discrimination, correlation matrix, and principal component axis loading were applied. During processing the stepwise discrimination, no variable (absorbance (Log(1/*R*)) was included in the model. According to Wilk's lambda value, a first selected variable with the most contribution was used for the classified model. The variable with the least contribution to the classified model had been removed in every step until no variable would be added or removed from the classified model. Matlab (2004) software used to write stepwise discrimination programs. The 69 most effective wavelengths selected by stepwise discrimination were 1100, 1120, 1124, 1152, 1160, 1164, 1168, 1208, 1240, 1244, 1248, 1288, 1292, 1296, 1464, 1500, 1520, 1584, 1616, 1672, 1676, 1696, 1728, 1740, 1768, 1776, 1784, 1812, 1816, 1844, 1856, 1888, 1892, 1896, 1900, 1908, 1920, 1924, 1936, 1944, 1948, 1964, 1968, 1988, 2000, 2004, 2020, 2040, 2044, 2060, 2072, 2084, 2092, 2104, 2116, 2124, 2136, 2160, 2168, 2200, 2232, 2300, 2332, 2344, 2348, 2356, 2396, 2412, and 2432 nm. Corresponding to the above list wavelength, the variable (absorbance (Log(1/*R*)) was used to establish models 3 and 4. Model 3 comprised 69 variables input into a discriminant analysis program. Model 4 comprised 69 variables input into a backpropagation neural network with 40 nodes in the first hidden layer and 20 nodes in the second hidden layer, with 50,000 iterations, an RMS error of 0.1, and a correlation coefficient of 0.91.

After principal component processing, each variable had a loading value between –1 and 1 on each principal component. The loading value reflected both how much the variable contributed to the principal component and how well the principal component was taken into account of the variable variation over the data points (Unscramber, CAMO ASA, Oslo, Norway). Then the absolute value for each loading value of each variable was rearranged for each wavelength by loading value from high to low. As shown in Fig. 1, the variance included 98.4% of the information in the first and second principal component axes. Using the largest absolute loading values from the first and second principal component axes, the 69 most effective wavelengths were 2496, 2492, 2500, 2488, 2484, 2480, 2476, 1104, 1100, 1116, 1108, 1112, 2472, 1120, 1124, 2468, 1128, 1132, 2464, 1136, 2460, 1140, 2456, 1144, 2452, 1148, 2448, 1152, 2444, 1156, 2440, 1160, 1164, 2436, 1168, 1292, 1296, 1300, 1288, 1284, 1304, 2432, 1172, 1280, 1276, 1308, 1272, 1268, 1264, 1260, 1312, 1256, 1176, 1252, 1248, 1316, 1244, 2428, 1240, 1320, 1180, 1236, 1324, 1232, 1184, 2424, 1228, 1328, and 1188 nm. Corresponding to

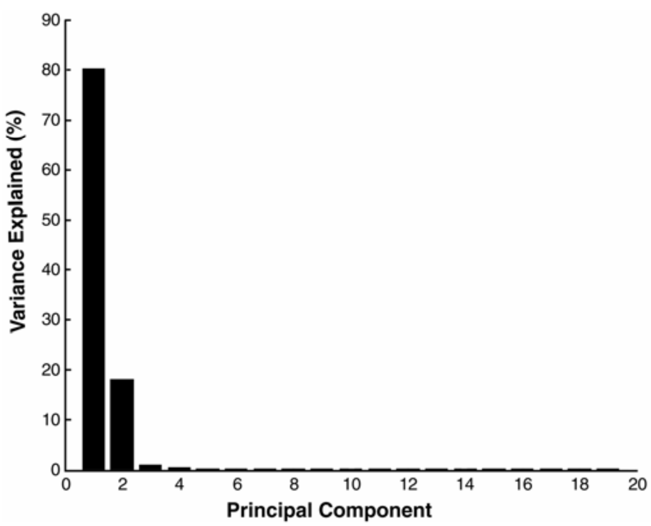

Fig. 1. Variance explained of principal components.

TABLE I
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 1[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	97 (100%)	84 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taichung Sen 10	0 (0%)	0 (0%)	96 (100%)	82 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	94 (97.9%)	80 (96.4%)	0 (0%)	0 (0%)	2 (2.1%)	3 (3.6%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (3.7%)	98 (100%)	79 (96.3%)	0 (0%)	0 (0%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.2%)	1 (1.0%)	1 (1.2%)	96 (99.0%)	80 (97.6%)

[a] Sample size with rate % (in parentheses).

TABLE II
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 2[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	94 (96.9%)	84 (100%)	3 (3.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taichung Sen 10	0 (0%)	0 (0%)	96 (100%)	82 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	96 (100%)	78 (94.0%)	0 (0%)	5 (6.0%)	0 (0%)	0 (0%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	13 (15.9%)	95 (96.9%)	68 (82.9%)	3 (3.1%)	1 (1.2%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.1%)	1 (1.2%)	4 (4.1%)	11 (13.4%)	92 (94.8%)	70 (85.4%)

[a] Sample size with rate % (in parentheses).

the above list wavelength, the variable (absorbance (Log(1/*R*)) was used to establish models 5 and 6. Model 5 comprised 69 variables input into a discriminant analysis program. Model 6 was found by inputting 69 variables into a backpropagation neural network for training with 40 nodes in the first hidden layer and 20 nodes in the second hidden layer, with 50,000 iterations, an RMS error of 0.19, and a correlation coefficient of 0.67.

In the manner of Paliwal et al (2003), the two-dimensional correlation coefficient matrix of each variable ($m \times n$ matrix, where m and n are sample size and number of variables, respectively) was found. Taking the absolute of each element and calculating the average of each column, this reduced the two-dimensional correlation coefficient matrixes to a one-dimensional correlation coefficient vector and rearranged each variable by correlation coefficient from low to high. Matlab software was used to write correlation coefficient matrix programs. The 69 most effective wavelengths selected by correlation coefficient matrix were 1100, 1104, 1108, 1112, 1116, 1120, 1124, 1128, 1132, 1136, 1140, 1144, 1148, 1152, 1156, 1160, 1164, 1168, 1172, 1292, 1296, 1300, 1288, 1304, 1284, 1280, 1308, 1176, 1276, 1268, 1272, 1260, 1264, 1312, 1256, 1252, 1248, 1316, 1180, 1244, 1240, 1320, 2496, 2500, 1236, 2492, 1184, 2488, 1232, 1324, 1228, 2484, 1188, 1224, 1328, 2480, 1192, 1220, 2476, 1196, 1216, 1332, 1200, 1212,1204, 1208, 2472, 1336, and 2468 nm. Corresponding to the above list wavelength, the variable (absorbance (Log(1/*R*)) was used to establish models 7 and 8. Model 7 comprised 69 variables input into a discriminant analysis program. Model 8 comprised 69 variables input into a backpropagation neural network for training with 50 nodes in the first hidden layer, 30 nodes in the second hidden layer, an RMS error of 0.12, correlation coefficient of 0.69, and with 50,000 iterations.

RESULTS AND DISCUSSIONS

Table I showed the validation and classification rates and sample size of model 1. With 484 validation samples, only two Tainung 67 and one Taikeng 9 were not validated correctly, and the average validation rate was 99.4% with 0.93% standard deviation. Three Tainung 67, three Taikeng 8, and two Taikeng 9 of the classification set were not classified correctly, and the average validation and classification rates were 99.4% with 0.93% standard deviation and 98.1% with 1.8% standard deviation, respectively.

Table II shows the average validation and classification rates of model 2 were 97.7% with 2.2% standard deviation and 92.5% with 8.0% standard deviation, respectively. Using the *z*-test from Shen (1999), model 1 is better than model 2 for classification.

Table III showed the validation and classification rates of model 3. The average validation and classification rates were 99.6% with 0.6% standard deviation and 98.5% with 1.6% standard deviation, respectively. After performing the *z*-test, models 3 and 1 showed no significant difference for classification. However, only 19.7% variables in model 1 were used in model 3. By *z*-test, model 3 is better than model 2.

Table IV showed the average validation and classification rates of model 4 were 96.9% with 1.4% standard deviation and 85.5% with 17.8% standard deviation, respectively. Comparing model 4 with model 3, model 3 is better than model 4. Model 4 is also better than model 2.

Table V showed the validation and classification rates of model 5. The average validation and classification rates were 94.6% with 2.9% standard deviation and 72.0% with 38.7% standard deviation. Comparing the classification rate of model 5 with model 3, model 3 is better than model 5.

TABLE III
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 3[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	96 (99.0%)	83 (98.8%)	1 (1.0%)	1 (1.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taichung Sen 10	0 (0%)	0 (0%)	96 (100%)	82 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	96 (100%)	83 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.4%)	97 (99.0%)	80 (97.6%)	1 (1.0%)	0 (0%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (3.7%)	0 (0%)	0 (0%)	97 (100%)	79 (96.3%)

[a] Sample size with rate % (in parentheses).

TABLE IV
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 4[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	94 (96.9%)	82 (97.6%)	3 (3.1%)	1 (1.2%)	0 (0%)	0 (0%)	0 (0%)	1 (1.2%)	0 (0%)	0 (0%)
Taichung Sen 10	3 (3.1%)	2 (2.4%)	93 (96.9%)	80 (97.6%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	3 (3.6%)	95 (99.0%)	54 (65.1%)	0 (0%)	26 (31.3%)	1 (1.0%)	0 (0%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	93 (94.9%)	82 (100%)	5 (5.1%)	0 (0%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.0%)	0 (0%)	2 (2.16%)	27 (32.9%)	94 (96.9%)	55 (67.1%)

[a] Sample size with rate % (in parentheses).

TABLE V
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 5[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	87 (89.7%)	83 (98.8%)	10 (10.3%)	1 (1.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taichung Sen 10	2 (2.1%)	13 (15.9%)	93 (96.9%)	68 (82.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.0%)	1 (1.2%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	92 (95.8%)	74 (89.2%)	4 (4.17%)	6 (7.2%)	0 (0%)	3 (3.6%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	10 (12.2%)	94 (95.9%)	70 (85.4%)	4 (4.1%)	2 (2.4%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (3.1%)	25 (30.5%)	2 (2.1%)	54 (65.9%)	92 (94.9%)	3 (3.7%)

[a] Sample size with rate % (in parentheses).

Table VI showed the average validation and classification rates were 82.6% with 11.06% standard deviation and 72.2% with 33.7% standard deviation, respectively. Comparing the rates from *z*-test, models 6 and 5 showed no significant difference. Comparing models 6 and 4, model 4 shows better wavelength selection by stepwise discrimination because of labeling paddy rice cultivars in each sample during variable selection.

Table VII showed the average of validation and classification of model 7 were 94.0% with 4.9% standard deviation and 69.1% with 37.1% standard deviation, respectively. Model 7 and 5 showed no significant difference for classification.

In Table VIII, the average validation and classification rates of model 8 were 85.5% with 8.9% standard deviation and 60.6% with 41.46% standard deviation, respectively. By *z*-test, model 6 is better than model 8 in classification. With significant difference, the priority of wavelength selection method was recommended in the following order: stepwise discrimination, correlation, and then loading values.

CONCLUSIONS

With paddy rice reflectance spectra of 1100 to 2500 nm in 3-nm steps, a total of 351 variables were used to develop model 1 with an average classification rate of 98.1%. Using the same 351 variables for a two hidden layers backpropagation neural network were used to establish model 2 with an average classification rate of 92.5%.

Sixty-nine variables selected by the stepwise discrimination method were used to develop the discriminant analysis model 3 with an average classification rate of 98.5%. The same 69 variables were input into a back-propagation neural network for training to produce model 4 with an average classification rate of 85.5%. Selecting 69 variables by the correlation from low to high were used to develop the discriminant analysis model 5 with an average classification rate of 72.0%. Using the same 69 variables in a backpropagation neural network to develop model 6, the average classification rate was 72.3%. The 69 variables selected by the loading values on the first and second principal components were used to develop the discriminant analysis model 7 and then produced an average classification rate of 69.1%. Using the same 69 variables as model 7 in the backpropagation neural network, model 8 had an average classification rate of 60.6%. The wavelength selection methods for near-infrared classification model were recommended in the priority order from stepwise discrimination, correlation, and then loading value. In classifying five popular paddy rice cultivars in Taiwan by near-infrared spectra, the discriminant method was more stable than the backpropagation neural network and required fewer variables. Model 3 with 98.5% classification rate is the best model because it only required 69 variables. Comparing Liu's morphological and color features model of 99.8% classification rate with model 3 with 98.5% classification rate, there is no significant difference, but model 3 may require annual recalibration by inputting about 80 samples of new crop year or ≈8% of the three-crop-year sample size in training set. This result may imply that annual changes of morphological and color features are less affected by climate, soil, and fertilizer, etc.

ACKNOWLEDGMENTS

We wish to express our appreciation for the financial support received from the National Science Council of Taiwan, R.O.C. (project No: NSC89-2313-B-002-241 and NSC90-2313-B-002-327). We also wish to

TABLE VI
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 6[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	81 (83.5%)	51 (60.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20 (23.8%)	0 (0%)	0 (0%)
Taichung Sen 10	16 (16.7%)	4 (4.9%)	70 (72.9%)	71 (86.6%)	0 (0%)	0 (0%)	0 (0%)	7 (8.5%)	10 (10.4%)	0 (0%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	67 (69.8%)	81 (97.6%)	19 (19.8%)	2 (2.4%)	10 (10.4%)	0 (0%)
Taikeng 8	1 (1.0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.4%)	92 (93.9%)	80 (97.6%)	5 (5.1%)	0 (0%)
Taikeng 9	0 (0%)	0 (0%)	70 (72.9%)	71 (86.6%)	6 (6.2%)	1 (1.2%)	1 (1.0%)	66 (80.5%)	90 (92.8%)	15 (18.3%)

[a] Sample size with rate % (in parentheses).

TABLE VII
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 7[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	83 (85.6%)	84 (100%)	14 (14.4%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Taichung Sen 10	2 (2.1%)	18 (22.0%)	94 (97.9%)	63 (76.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.2%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	91 (94.8%)	76 (91.6%)	3 (3.1%)	4 (4.8%)	2 (2.1%)	3 (3.6%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	20 (24.4%)	95 (96.9%)	58 (70.7%)	3 (3.1%)	4 (4.9%)
Taikeng 9	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (2.1%)	19 (23.2%)	3 (3.1%)	58 (70.7%)	92 (94.9%)	5 (6.1%)

[a] Sample size with rate % (in parentheses).

TABLE VIII
Validation (Val) and Classification (Cal) Rates (%) and Sample Size of Model 8[a]

	Tainung Sen 20		Taichung Sen 10		Tainung 67		Taikeng 8		Taikeng 9	
	Val	Cal	Val	Cal	Val	Cal	Val	Cal	Val	Cal
Tainung Sen 20	89 (91.8%)	30 (35.7%)	8 (8.3%)	26 (31.0%)	0 (0%)	0 (0%)	0 (0%)	28 (33.3%)	0 (0%)	0 (0%)
Taichung Sen 10	14 (14.6%)	3 (3.7%)	80 (83.3%)	65 (79.3%)	0 (0%)	0 (0%)	0 (0%)	13 (15.9%)	2 (2.1%)	1 (1.2%)
Tainung 67	0 (0%)	0 (0%)	0 (0%)	0 (0%)	69 (71.9%)	76 (91.6%)	14 (14.6%)	7 (8.4%)	13 (13.5%)	0 (0%)
Taikeng 8	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (3.7%)	93 (94.9%)	79 (96.3%)	5 (5.1%)	0 (0%)
Taikeng 9	1 (1.1%)	0 (0%)	2 (2.1%)	0 (0%)	9 (9.3%)	0 (0%)	0 (0%)	41 (50%)	83 (85.6%)	0 (0%)

[a] Sample size with rate % (in parentheses).

convey our thanks and appreciation to Yud-Ren Chen, research leader of ISL, USDA, ARS, BA, and to Chou Jui-Jen and Lin Ta-Te, Dept. of Bio-Industrial Mechatronics Engineering, National Taiwan University, for their helpful advice.

LITERATURE CITED

Chang, H., Shaw, J., Lee, S., and Hong, M. 2000. Classifying varieties of paddy rice by near-infrared and image processing technique. J. Agric. Mach. 9:1-16.

Corbella, E., and Cozzolino, D. 2005. The use of visible and near-infrared spectroscopy to classify the floral origin of honey samples produced in Uruguay. J. NIRS 13:63-68.

Delwiche, S. R., Chen, Y. R., and Hruschka, W. R. 1995. Differentiation of hard red wheat by near-infrared analysis of bulk samples. Cereal Chem. 72:243-247.

Liu, M., and Shaw, J. 1995. The detection of rice moisture and protein content with near-infrared spectrophotometer. J. Agric. Mach. 4:1-14.

Li, W., and Shaw, J. 1996. Data processing affecting the NIR calibration curves of major constituents of rough rice taste. J. Agric. Mach. 5:19-34.

Li, W. and Shaw, J. 1997. Determining the fat acidity of rough rice by near-infrared reflectance spectroscopy. Cereal Chem. 74:556-560.

Liu, C. C., Shaw, J. T., Poong, K. Y., Hong, M. C., and Shen, M. L. 2005. Classifying paddy rice by morphological and color features using machine vision. Cereal Chem. 82:649-653.

Paliwal, J., Visen, N. S., Jayas, D. S., and White, N. D. G. 2003. Comparison neural network and a non-parametric classifier for grain kernel identification. Biosys. Eng. 85:405-413.

Peng, Z. Y. 2000. SAS and statistical analysis (in Chinese). 2nd Ed. Pages 41-2. Ru-lin Publisher: Taipei, Taiwan.

Rencher, C. A. 1995. Discriminant analysis: Description of group separation. Pages 296-325 In: Methods of Multivariate Analysis. John Wiley and Sons: New York.

Rittiron, R., Saranwong, S., and Kawano, S. 2005. Detection of variety contamination in milled Japanese rice using a single kernel near infrared technique in transmittance mode. J. NIRS 13:9-25.

Shen, M. 1999. Introduction to statistics (in Chinese), 3rd Ed. Pages 177-178. Chiu-chow Publisher: Taipei, Taiwan.

Song, H., Delwiche, S. R., and Chen, Y. R. 1995. Neural network classification of wheat using single kernel near-infrared transmittance spectra. Opt. Eng. 34:2927-2934.

Velleman, P. F., and Welsch, R. E. 1981. Efficient computing of regression diagnostics. Am. Statist. 35:234-242.

Wang, D., Ram, M. S., and Dowell, F. E. 2002. Classification of damaged soybean seeds using neat infrared spectroscopy. Trans. ASAE 45:1943-1948.

[Received August 10, 2005. Accepted February 6, 2006.]