

CLASSIFICATION OF DAMAGED SOYBEAN SEEDS USING NEAR-INFRARED SPECTROSCOPY

D. Wang, M. S. Ram, F. E. Dowell

ABSTRACT. *Damage is an important quality factor for grading, marketing, and end use of soybean. Seed damage can be caused by weather, fungi, insects, artificial drying, and by mechanical damage during harvest, transportation, storage, and handling. The current visual method for identifying damaged soybean seeds is based on discoloration and is subjective. The objective of this research was to classify sound and damaged soybean seeds and discriminate among various types of damage using NIR spectroscopy. A diode-array NIR spectrometer, which measured reflectance spectra ($\log[1/R]$) from 400 to 1,700 nm, was used to collect single-seed spectra. Partial least square (PLS) models and neural network models were developed to classify sound and damaged seeds. For PLS models, the NIR wavelength region of 490 to 1,690 nm provided the highest classification accuracy for both cross-validation of the calibration sample set and prediction of the validation sample set. Classification accuracy of sound and damaged soybean seeds was higher than 99% when using a two-class model. The classification accuracies of sound seeds and those damaged by weather, frost, sprout, heat, and mold were 90%, 61%, 72%, 54%, 84%, and 86%, respectively, when using a six-class model. Neural network models yielded higher classification accuracy than PLS models. The classification accuracies of the validation sample set were 100%, 98%, 97%, 64%, 97%, and 83% for sound seeds and those damaged by weather, frost, sprout, heat, and mold, respectively, for the neural network model. The optimum parameters of the neural network model were learning rate of 0.7 and momentum of 0.6.*

Keywords. *Damage, Near-infrared spectroscopy, Neural networks, Soybean seeds.*

Soybean is one of the major oil seed crops in the U.S., especially in the Midwest. U.S. farmers produce more than 60% of the world's total soybean production (Soyatech, 2001). More than one-third of U.S. soybean production is exported to foreign markets. The production and marketing of soybean are major components of the U.S. agricultural economy. Assessment of soybean quality is an important task for government agencies, such as the Grain Inspection, Packers, and Stockyard Administration (GIPSA), local elevators, and soybean industries.

The U.S. standard used to evaluate soybean quality is based on test weight, damaged kernels, foreign materials, and discolored kernels (USDA, 1994). The maximum limits for damaged soybeans are $\geq 1\%$, 2%, 5%, and 10% for U.S. Grade No. 1, 2, 3, and 4, respectively. The maximum limits on damaged soybean seeds are very restricted, especially for U.S. Grade No. 1 and 2 soybeans. Currently, GIPSA and local elevators visually determine damaged soybean seeds from sound seeds based on inspection of physical quality attrib-

utes, such as color, shape, and size. Although reference slides are available to assist inspectors in making classification judgments, distinguishing damaged seeds from sound seeds is still subjective. The variability caused by this subjectivity is difficult to document, but it may affect grade and market value of soybeans. In addition, the various types of damage have different effects on soybean seed quality and end use. Correct classification of damage type can provide useful information for the end use of the soybeans and can therefore benefit both soybean producers and industries. Grading requires repetition, adaptability, and objectivity, which for human inspectors may change over a long period of grading. Thus, an objective grading system is needed to improve marketing accuracy.

Currently, the methods most studied for objective classification of sound and damaged soybean seeds are based on machine vision, limited to a few damage categories, and restricted to classifying sound or fungal-damaged soybean seeds. Machine vision is the application of electronic imaging to enable the visual inspection of an object or a scene (Paulsen, 1990). Wigger et al. (1988) developed a color image processing system to classify soybean seeds into healthy and symptomatic seeds, focusing on fungal-damaged seeds. The fungal-damaged seeds were correctly determined in 77% to 91% of the soybean seeds tested. Casady et al. (1992) developed an image pattern classification program to discriminate healthy seeds and fungal-damaged seeds, with average classification accuracy of 77% to 91%. Ahmad et al. (1999) used a color classifier for classifying fungal-damaged soybean seeds. The color analysis showed differences between the asymptomatic and symptomatic seeds. However, color alone did not adequately describe some differences among symptoms. The average

Article was submitted for review in June 2002; approved for publication by the Food & Process Engineering Institute Division of ASAE in October 2002. Contribution No. 02-393-J from the Kansas Agricultural Experiment Station.

The authors are **Donghai Wang, ASAE Member**, Assistant Professor, Department of Biological and Agricultural Engineering, Kansas State University, Manhattan, Kansas; and **M. S. Ram, Chemist**, and **Floyd E. Dowell, ASAE Member Engineer**, Research Leader, USDA-ARS, Grain Marketing and Production Research Center, Manhattan, Kansas. **Corresponding author:** Donghai Wang, Department of Biological and Agricultural Engineering, Kansas State University, Manhattan, KS 66506; phone: 785-532-2919; fax: 785-532-5825; e-mail: dwang@ksu.edu.

classification accuracy was 88% for the asymptomatic and symptomatic seeds. The classification accuracies for *Phomopsis lonicolla*, *Alternaria*, *Fusarium graminearum*, and *Cercospora kikuchii* were 45%, 30%, 62%, and 83%, respectively. Shatadal and Tan (1998) developed a four-class neural network model based on color image analysis for classification of sound, heat-damaged, frost-damaged, and stinkbug-damaged soybean seeds. The classification accuracies were 99.6%, 95%, 90%, and 50.6% for sound, heat-damaged, frost-damaged, and stinkbug-damaged soybean seeds, respectively.

Image analysis can provide a means of extracting useful information from an image for quality measurement of grains and oil seeds. However, machine vision cannot provide specific information related to chemical composition because machine vision is limited to the visible region of the spectrum. Near-infrared (NIR) spectroscopy can be used for measurement of both physical and chemical properties. Therefore, NIR spectroscopy may increase the classification accuracy of sound and damaged soybean seeds. The objective of this research was to classify sound and damaged soybean seeds and discriminate among various types of damage using NIR spectroscopy.

MATERIALS AND METHODS

MATERIALS

Market-channel soybean samples (not variety specific) were obtained from the GIPSA Federal Grain Inspection Service (FGIS) and the Department of Agronomy, Kansas State University. Sound and damaged seeds were manually classified by experienced grain inspectors. Seeds of six categories (sound, weather-damaged, frost-damaged, sprout-damaged, heat-damaged, and mold-damaged) were used for this study. Characteristics of each category and the number of seeds used for this study are listed in table 1.

KERNEL COLOR MEASUREMENT

Reflectance spectra from 490 to 750 nm were transferred into $L^*a^*b^*$ color space using Grams/32 (Galactic Industries, Salem, N.H.) software and were used for soybean seed color determination. In the $L^*a^*b^*$ color space, L^* varies from 0 (black) to 100 (perfect white); a^* ranges from -100 to 100 and measures green when negative and red when positive; and b^* varies from -100 to 100 and is a measure of blue when negative and yellow when positive.

Table 1. Characteristics of sound and damaged soybean samples and the number of soybean seeds used for classification of sound and damaged soybeans.

Type of Damage	No. of Seeds	Descriptions of Damage
Sound seeds	700	Soybeans with natural yellowish color.
Weather-damaged	200	Soybeans with discolored seed coat.
Frost-damaged	200	Soybeans are discolored green in cross-section.
Heat-damaged	200	Soybeans are materially discolored and damaged by heat.
Sprout-damaged	100	Soybeans are immature and have a thin, flat, and wrinkled appearance.
Surface mold damage (downy mildew)	200	Soybeans with milky white or grayish crusty growth.

NIR SPECTRA COLLECTION

A diode-array NIR spectrometer (DA7000, Perten Instruments, Springfield, Ill.) was used to collect single seed spectra. The spectrometer measures reflectance from 400 to 1700 nm using an array of silicon sensors for spectrum region from 400 to 950 nm at 7-nm bandwidth and indium-gallium arsenide sensors for spectrum region from 950 to 1700 nm at 11-nm bandwidth. All data is then interpolated to 5-nm intervals. The diode-array NIR spectrometer collected spectra at a rate of 30/s. Single soybean seeds were placed in a black V-shaped trough (12 mm long, 10 mm wide, and 5 mm deep) and illuminated with halogen light via a fiber bundle (8 mm diameter) positioned 13 mm from the top of the trough and oriented 45° from vertical. A 2-mm reflectance probe, oriented vertically 9.5 mm from the top of the trough, carried the reflected energy to a spectrometer. The procedures included collecting a baseline, collecting eight spectra from each seed, and averaging the eight spectra for each seed. A total of 700 sound seeds and a total of 900 seeds damaged by weather, frost, sprout, heat, or mold were measured. A spectrum of the empty trough was measured as a reference before seed measurement, and again after every 100 seeds.

PARTIAL LEAST SQUARES

Partial least squares (PLS) software (Galactic Industries, Salem, N.H.) was used to develop two-class and six-class models for classification of sound and damaged soybean seeds. PLS is a multivariate data analysis technique designed to handle intercorrelated regressors. For two-classification models, soybean seeds first were separated equally into calibration and validation sets, based on even and odd numbers. Sound and damaged seeds were assigned constant values of 1.0 and 2.0, respectively. A seed was considered to be correctly categorized if the predicted value lay on the same side of the midpoint of the assigned values. For six-class models, sound seeds and weather, frost, sprout, heat, and mold damaged seeds were assigned constant values of 1.0, 2.0, 3.0, 4.0, 5.0, and 6.0, respectively. The model performance is reported as the cross-validation of each calibration sample set and prediction of validation sample sets. The number of PLS factors used was the minimum required to give the best classification results.

NEURAL NETWORKS

The NeuralWorks Professional II/Plus software package (NeuralWare, Inc., Pittsburgh, Pa.) was used to develop neural network models for classification of sound and damaged soybean seeds based on back-propagation networks. Back-propagation uses a learning process to minimize the global error of the system by modifying node weights. The weight increment or decrement is achieved by using the gradient descent rule. The network is trained by initially selecting the weights at random and then presenting all training data repeatedly. The weights are adjusted after every trial using external information specifying the correct result until the weights converge and the errors are reduced to acceptable values. A complete discussion of back-propagation network theory is given by Hecht-Nielsen (1989). Visible (490–750 nm), NIR (750–1690 nm), and full wavelength (490–1690 nm) regions were used as neural network inputs. For each classification experiment, two types of neural networks (with and without a hidden layer) were

Table 2. Classification accuracy (%) of sound and damaged soybean seeds using two-class partial least squares (PLS) models.

Spectral Region	Calibration Results ^[a]				Validation Results ^[b]		
	F ^[c]	Sound	Damaged	Average	Sound	Damaged	Average
490–750 nm	6	98.8	98.4	98.6	98.2	97.8	98.0
750–1690 nm	10	100	100	100	99.7	99.7	99.7
490–1690 nm	10	99.7	99.3	99.5	99.7	99.6	99.6

[a] Total number of soybean seeds in the calibration sample set = 800.

[b] Total number of soybean seeds in the validation sample set = 800.

[c] F = number of PLS regression factors.

tested. Six-class neural network models were developed. The network with the highest validation accuracy was recorded as the best model.

RESULTS AND DISCUSSION

CLASSIFICATION OF SOUND AND DAMAGED SOYBEANS USING PLS

Classification results of cross-validation of the calibration sample sets and prediction of the validation sample sets using two-class PLS models are summarized in table 2. The NIR wavelength region of 750–1690 nm and visible/NIR region of 490–1690 nm gave the highest percentage of correct classification for both cross-validation and prediction (>99.5%). The use of the visible wavelength alone resulted in the poorest classification performance. Sound seeds and damaged seeds have both different physical properties and different chemical compositions. The use of the NIR region appears to have added important information. Sound soybean seeds are usually a natural yellowish color with an intact seed coat. In contrast, damaged soybean seeds are usually discolored and have fissures in the seed coat, resulting in wrinkled, cracked, and crusty seed coats with less roundness. Heat damage can cause discoloration and can also cause protein denaturation. Figure 1 shows the average spectra of sound and damaged soybean seeds. When compared to sound soybean seeds, the energy absorption of damaged soybean seeds was higher in the visible region and lower in the NIR region. The differences in energy absorption indicate differences in color and chemical composition between damaged and sound soybean seeds. The greater energy absorption of damaged soybean seeds in the visible region indicates that the color of damaged soybean seeds was

darker than sound soybean seeds. This result was expected because damaged seeds were usually accompanied by discoloration.

The color differences between damaged and sound seeds were further shown by *L*, *a*, and *b* values (table 3). In general, sound soybean seeds had higher average *L*, *a*, and *b* values than damaged soybean seeds. This response indicates that damaged soybean seeds are darker and more yellow than sound soybean seeds. However, there were no significant differences in *L* and *b* values between sound seeds and heat-damaged and sprout-damaged soybean seeds. Hence, there was very little discrimination information in the visible region of 490–750 nm, which resulted in relatively poor classification performance.

The peaks and valleys of the beta coefficients curve from the PLS models show the significant differences in energy absorption between the sound and damaged soybean seeds (fig. 2). The peaks around 515, 580, 640, 690, and 725 nm are related to color difference between sound and damaged soybean seeds. The peaks around 970 and 1515 nm are related to NH bonds, representing protein content (Shenk et al., 2001), and the peaks around 1215, 1345, and 1600 nm are related to CH bonds, representing fiber content (Barton and Burdick, 1979; Shenk et al., 2001). The damaged soybean seeds may have less protein and fiber contents than sound soybean seeds because damaged soybean seeds have less weight than sound soybean seeds (Mbuvi et al., 1989; Sinclair, 1995). The peak around 1410 nm is related to OH bonds, representing oil content (Shenk et al., 2001). Some peaks and valleys in the beta coefficient curve may represent interactions of moisture, starch, protein, oil, and cellulose caused by damage to the seeds.

Table 4 shows classification results of cross-validation of the calibration sample sets and prediction of the validation sample sets using six-class PLS models. Classification accuracy of cross-validation and the prediction of the validation sample sets were similar. The classification model

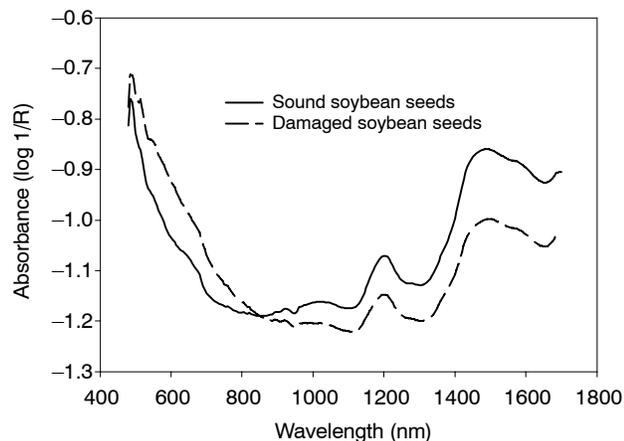


Figure 1. NIR absorption curves for sound and damaged soybean seeds.

Table 3. The color variations between sound and damaged soybean seeds measured as *L*, *a*, and *b* values in the *L*a***b** color space.**

Classification	<i>L</i>	<i>a</i>	<i>b</i>
Sound soybean	36.01 a ^[a]	6.89 c	9.52a
Damaged soybean			
Weather-damaged	26.30 c	5.77 d	6.40 b
Frost-damaged	33.44 ab	3.26 f	5.65 b
Sprout-damaged	33.59 a	9.15 a	8.93 a
Heat-damaged	33.74 a	7.82 b	9.08 a
Mold-damaged	30.58 b	4.99 e	5.88 b
Average	31.54	6.20	7.19
LSD	2.87	0.71	0.95

[a] Values within the same column followed by different letters are significantly different at *P* < 0.05.

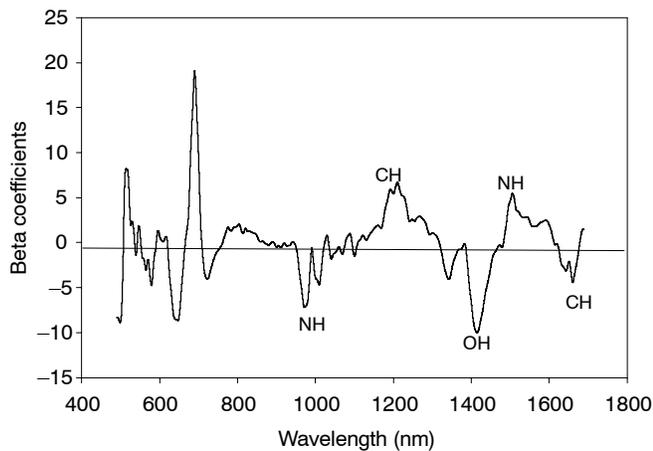


Figure 2. Beta coefficients curve of PLS model (10 PLS factors) for classification of sound and damaged soybean seeds.

using the full wavelength region (490–1690 nm) yielded the highest average classification accuracy for both calibration and validation sample sets (>74.5%). Classification accuracies were 90.2%, 61%, 72%, 54%, 84%, and 86%, corresponding to sound seeds and seeds damaged by weather, frost, sprout, heat, and mold, respectively. The visible (490–750 nm) and NIR (750–1690 nm) wavelength regions alone gave lower classification accuracies. Hence, each wavelength region contained valuable information representing the physical or chemical properties of the soybean seeds. The classification accuracy may indicate a non-linear relationship among the different types of damage. Sprout-damaged soybean seeds were the most difficult to classify (table 4). This result was probably because sprout-damaged soybean seeds had a color similar to sound and heat-damaged seeds.

CLASSIFICATION OF SOUND AND DAMAGED SOYBEANS USING NEURAL NETWORKS

Neural network classification models were developed using three wavelength regions: 490–750 nm with 53 input nodes, 750–1690 nm with 189 input nodes, and 490–1690 nm with 241 input nodes. Spectral data below 490 nm and above 1690 nm were discarded because of the noise in the signal. Six output nodes were used to classify soybean samples into six categories. The delta rule was used as the learning rule, which specifies how connection weights are changed during

the learning process. The sigmoid transfer function was chosen for the non-linear function that transfers the internally generated sum for each node to a potential output node.

Previous research showed no benefit from using more than one hidden layer (Bochereau et al., 1992; Dowell, 1994; Wang et al., 1999). Thus, only the differences between one hidden layer and no hidden layer were compared. The number of nodes in the hidden layer was varied among 0, 3, 6, and 25. For each of these, 150,000 learning cycles (1875 cycles/spectrum), momentum of 0.4, and learning rate of 0.5 were used for this comparison. Adding a hidden layer decreased classification accuracy for both the calibration and validation sample sets (table 5). Adding a hidden layer also strongly affected classification of sprout-damaged and mold-damaged categories. Increasing the number of nodes in the hidden layer from 3 to 25 increased the overall classification accuracy and the average accuracy for mold-damaged seeds, but the classification of sprout-damaged seeds was still poor. The best classification was obtained by using a neural network model without hidden layers. Thus, only the neural network models without the hidden layer will be discussed further.

The major parameters that affect the classification and training speed of the neural networks are the number of learning cycles (also called events), learning rate (η), and momentum (α). Neural network models with a wavelength region of 490–1690 nm and optimum learning cycles of 150,000 were used to find the optimum learning rate and momentum. Table 6 shows the effect of learning rate and momentum on average classification accuracy of the calibration and validation sample sets. The results show that when the momentum was kept constant, the highest learning rate resulted in the highest classification accuracy. Learning rate determines how much of the error to propagate back into the preceding nodes, and it affects the speed of convergence of the network. High learning rates may need fewer learning cycles to achieve the required error level and thus use less training time. The classification accuracy increased as momentum and learning rate increased. At a constant learning rate, a momentum of 0.6 yielded the highest classification accuracy. Likewise, at constant momentum, a learning rate of 0.7 gave the best classification results. Therefore, a momentum of 0.6 and learning rate of 0.7 were used for network training.

The performance of neural network models for three different wavelength regions (490–750, 750–1690, and

Table 4. Classification accuracy of sound and damaged soybean seeds using six-class partial least squares (PLS) models.^[a]

Sample Sets	Classification Accuracy (%)						Average
	Sound	Weather Damaged	Frost Damaged	Sprout Damaged	Heat Damaged	Mold Damaged	
Calibration set ^[b]							
490–750 nm	85.0	64.0	53.0	80.0	88.0	68.0	70.0
490–1690 nm	88.0	62.0	72.0	56.0	86.0	87.0	75.2
750–1690 nm	70.5	57.0	60.0	50.0	77.0	80.0	65.8
Validation set ^[c]							
490–750 nm	86.5	67.0	45.0	76.0	84.0	77.0	72.5
490–1690 nm	90.2	61.0	72.0	54.0	84.0	86.0	74.5
750–1690 nm	70.2	79.0	71.0	40.0	55.0	81.0	66.0

^[a] Number of PLS regression factors = 10.

^[b] Total number of soybean seeds in the calibration sample set = 650.

^[c] Total number of soybean seeds in the validation sample set = 800.

Table 5. Effect of number of hidden layer nodes on classification accuracy of sound and damaged soybean seeds using six-class neural network models.^[a]

No. of Hidden Layer Nodes	Classification Accuracy (%)						
	Sound	Weather Damaged	Frost Damaged	Sprout Damaged	Heat Damaged	Mold Damaged	Average
Calibration set ^[b]							
0	99.7	97.0	96.0	66.0	94.0	80.0	88.8
3	99.7	100	99.0	0	0	5.0	50.6
6	99.7	96.0	93.0	0	95.0	69.0	75.5
25	99.4	97.0	92.0	10.0	95.0	90.0	80.6
Validation set ^[c]							
0	100	98.0	95.0	54.0	94.0	57.0	83.0
3	100	99.0	96.0	0	7.0	6.0	51.3
6	100	96.0	91.0	0	92.0	63.0	73.7
25	99.7	93.0	92.0	6.0	95.0	87.0	78.8

^[a] Parameters used: wavelength range = 490–1690 nm; momentum = 0.4; learning rate = 0.5; and learning cycles = 150,000.

^[b] Total number of soybean seeds in the calibration sample set = 800.

^[c] Total number of soybean seeds in the validation sample set = 800.

Table 6. Effect of learning rate and momentum on average classification accuracy of calibration and validation sample sets using six-class neural network models.^[a]

Momentum	Learning Rate	Classification Accuracy (%)						
		Sound	Weather Damaged	Frost Damaged	Sprout Damaged	Heat Damaged	Mold Damaged	Average
0.4	0.5	100	97.5	95.5	60.0	94.0	68.5	85.9
0.4	0.6	100	98.0	95.5	62.0	94.5	72.5	87.1
0.4	0.7	100	98.0	97.0	69.0	96.0	78.0	89.7
0.5	0.5	100	97.5	95.5	62.0	94.0	66.5	85.9
0.5	0.6	100	98.0	96.5	66.0	94.5	76.0	88.5
0.5	0.7	100	97.5	97.5	73.0	97.0	83.5	91.4
0.6	0.5	100	98.0	95.5	60.0	94.5	73.0	86.8
0.6	0.6	100	98.0	97.0	67.0	95.0	79.0	89.3
0.6	0.7	100	97.5	98.0	74.0	97.0	85.0	91.9

^[a] Parameters used: wavelength range = 490–1690 nm; learning cycles = 150,000; total number of soybean seeds in the calibration and validation sample sets = 800 in each set.

Table 7. Effect of wavelength region on classification accuracy of sound and damaged soybean seeds using six-class neural network models.^[a]

Sample Sets	Classification Accuracy (%)						
	Sound	Weather Damaged	Frost Damaged	Sprout Damaged	Heat Damaged	Mold Damaged	Average
Calibration set ^[b]							
490–750 nm	99.0	91.0	94.0	6.0	75.0	69.0	72.3
490–1700 nm	100	98.0	98.0	82.0	96.0	84.0	93.0
750–1690 nm	100	93.0	44.0	34.0	95.0	49.0	69.2
Validation set ^[c]							
490–750 nm	99.4	86.0	91.0	0	60.0	64.0	66.7
490–1690 nm	100	98.0	97.0	64.0	97.0	83.0	89.8
750–1690 nm	100	93.0	35.0	34.0	97.0	43.0	67.0
PLS 490–1690 nm	90.2	61.0	72.0	54.0	84.0	86.0	74.5

^[a] Parameters used: wavelength range = 490–1690 nm; momentum = 0.6; learning rate = 0.7; and learning cycles = 150,000.

^[b] Total number of soybean seeds in the calibration sample set = 800.

^[c] Total number of soybean seeds in the validation sample set = 800.

490–1690 nm) is summarized in table 7. In general, classification accuracy for the calibration sample sets was higher than for the validation sample sets. The neural network using the 490–1690 nm wavelength region yielded the highest classification accuracy. More than 90% of sound, weather-damaged, frost-damaged, and heat-damaged soybean seeds were correctly classified for the calibration and validation sample sets, respectively. Sprout-damaged and

mold-damaged seeds were the most difficult to classify. However, 74% and 85% of sprout-damaged and mold-damaged soybeans were correctly classified when the 490–1690 nm wavelength region was used in the neural network model. Combining the effects of learning rate, momentum, learning cycles, and wavelength region on neural network performance, the optimum parameters of the neural network model were momentum = 0.6, learning rate = 0.7, learning

Table 8. Calibration and validation results using optimum six-class neural network models for classification of sound and damaged soybean seeds (490–1690 nm).

Type of Damage	Sample Set ^[a]	Classification Accuracy (%)					
		Sound	Weather Damaged	Frost Damaged	Sprout Damaged	Heat Damaged	Mold Damaged
Sound	Calibration	99.7	—	—	—	—	0.3
	Validation	100	—	—	—	—	—
Weather damaged	Calibration	0	97.0	1.0	2.0	—	—
	Validation	—	98.0	2.0	—	—	—
Frost damaged	Calibration	1.0	—	99.0	—	—	—
	Validation	2.0	—	97.0	1.0	—	—
Sprout damaged	Calibration	4.0	8.0	—	84.0	2.0	2.0
	Validation	14.0	10.0	—	64.0	6.0	6.0
Heat damaged	Calibration	—	2.0	—	1.0	97.0	—
	Validation	—	2.0	—	1.0	97.0	—
Mold damaged	Calibration	10.0	—	2.0	—	1.0	87.0
	Validation	13.0	—	4.0	—	—	83.0

[a] Total number of soybean seeds in the calibration and validation sample sets = 800 in each set.

cycles = 150,000, and wavelength region = 490–1690 nm with no hidden layer.

Table 8 shows the performance of the optimum neural network model on the validation sample set corresponding to six different seed categories. The classification accuracies of sound, weather-damaged, frost-damaged, and heat-damaged soybean seeds were higher than 97.0%. Sound soybean seeds had 100% classification accuracy. The sprout-damaged soybean seeds were the most difficult to classify (64%). About 36% of sprout-damaged soybean seeds were randomly misclassified into sound, weather-damaged, heat-damaged, and mold-damaged soybean categories. Mold-damaged soybean seeds were the second most difficult to classify. About 17% of mold-damaged soybean seeds were misclassified as sound or frost-damaged soybean seeds.

CONCLUSIONS

Sound and damaged soybean seeds can be easily classified using PLS models. The highest classification accuracy was more than 99.5% when either the wavelength regions of 750–1690 or 490–1690 nm were used under a two-class model. The classification of six categories of sound and damaged soybean seeds was significantly lower than two-class classification. The classification accuracies of the validation sample set were 90%, 61%, 72%, 54%, 84%, and 86%, corresponding to sound seeds and seeds damaged by weather, frost, sprout, heat, and mold, respectively, when using a six-class model. Sprout-damaged soybean seeds were the most difficult to classify. The neural network model yielded higher classification accuracy than the PLS model for six-class classification. The average of correct classifications was 94.0% for the calibration sample set and 89.8% for the validation sample set. The classification accuracies of the validation sample set were 100%, 98%, 97%, 64%, 97%, and 83% corresponding to sound seeds and weather, frost, sprout, heat, and mold damaged seeds, respectively. The optimum parameters of the neural network model were momentum = 0.6, learning rate = 0.7, learning cycles = 150,000, and wavelength region = 490–1690 nm.

Results from this limited sample source suggest that NIR technology could be used to detect damaged soybean seeds. More research is needed with a larger number of sample

sources to confirm these results for applying this technology to a commercial grading system.

REFERENCE

- Ahmad, I. S., J. F. Reid, and M. R. Paulsen. 1999. Color classifier for symptomatic soybean seeds using image processing. *Plant Disease* 83(4): 320–327.
- Barton II, F. E., and D. Burdick. 1979. Preliminary study on the analysis of forages with a filter-type near-infrared reflectance spectrometer. *J. Agric. Food Chem.* 27(6): 1248–1252.
- Bochereau, L., P. Bourguine, and B. Palagos. 1992. A method for prediction by combining data analysis and neural networks: Application to prediction of apple quality using near-infrared spectra. *J. Agric. Eng. Res.* 51(3): 207–216.
- Casady, W. W., M. R. Paulsen, J. F. Reid, and J. B. Sinclair. 1992. A trainable algorithm for inspection of soybean seed quality. *Trans. ASAE* 35(6): 2027–2033.
- Dowell, F. E. 1994. Neural network parameter effects on object classification and wavelength selection. *Optical Eng.* 33(8): 2728–2732.
- Hecht-Nielsen, R. 1989. *Neural Computing*. New York, N.Y.: Addison-Wesley.
- Mbuvu, S. W., J. B. Litchfield, and J. B. Sinclair. 1989. Physical properties of soybean seeds damaged by fungi. *Trans. ASAE* 32(6): 2093–2096.
- Paulsen, M. R. 1990. Using machine vision to inspect oilseeds. *INFORM* 1(1): 50–55.
- Shatadal, P., and J. Tan. 1998. Identifying damaged soybeans by using color image analysis. ASAE Paper No. 986023. St. Joseph, Mich.: ASAE.
- Shenk, J. S., J. J. Workman, and M. O. Westerhaus. 2001. Application of NIR spectroscopy to agricultural products. In *Hand Book of Near-Infrared Analysis*, 419–474. D. A. Burnes and E. W. Ciurczak, eds. New York, N.Y.: Marcel Dekker.
- Sinclair, J. B. 1995. Reevaluation of grading standards and discounts for fungus-damaged soybean seeds. *JAOC* 72(12): 1415–1419.
- Soyatech. 2001. *Soya Bluebook*. Bar Harbor, Maine: Soyatech.
- USDA. 1994. *Grain Inspection Handbook*. Washington, D.C.: USDA.
- Wang, D., F. E. Dowell, and R. E. Lacey. 1999. Single wheat kernel color classification using neural networks. *Trans. ASAE* 42(1): 233–240.
- Wigger, W. D., M. R. Paulsen, J. B. Litchfield, and J. B. Sinclair. 1988. Classification of fungal-damaged soybeans using color-image processing. ASAE Paper No. 883053. St. Joseph, Mich.: ASAE.