

Classification of Fungal-Damaged Soybean Seeds Using Near-Infrared Spectroscopy[#]

D. Wang,^{1,*} F. E. Dowell,³ M. S. Ram,³ and W. T. Schapaugh²

¹Department of Biological and Agricultural Engineering and

²Department of Agronomy, Kansas State University, Manhattan, Kansas, USA

³Grain Marketing and Production Research Center, USDA-ARS,
Manhattan, Kansas, USA

ABSTRACT

Fungal damage has a devastating impact on soybean quality and end-use. The current visual method for identifying damaged soybean seeds is based on discoloration and is subjective. The objective of this research was to classify healthy and fungal-damaged soybean seeds and discriminate among various types of fungal damage using near-infrared (NIR) spectroscopy. A diode-array NIR spectrometer, which measured reflectance [$\log(1/R)$] from 400 to 1700 nm, was used to obtain spectra from single soybean seeds. Partial least square (PLS) and neural network models were developed to differentiate healthy and fungal-damaged seeds. The highest classification accuracy was more than 99% when the wavelength region of 490–1690 nm was used under a two-class PLS model. Neural network models yielded higher classification accuracy than the PLS models for five-class classification. The average of correct classifications was 93.5% for the calibration sample set and 94.6% for the validation sample set. Classification accuracies of the validation sample set were 100, 99, 84, 94, and 96%

[#]Contribution No. 03-163-J from the Kansas Agricultural Experiment Station. Mention of a trademark or proprietary product does not constitute a guarantee or warranty of the product by the U.S. Department of Agriculture and does not imply its approval to the exclusion of other products that also may be suitable.

*Correspondence: D. Wang, Department of Biological and Agricultural Engineering, Kansas State University, Manhattan, KS 66506, USA; Fax: 785-532-5825; E-mail: dwang@ksu.edu.

corresponding to healthy seeds, *Phomopsis*, *Cercospora kikuchii*, soybean mosaic virus (SMV), and downy mildew damaged seeds, respectively.

Key Words: Soybean seeds; Fungal-damage; Near-infrared spectroscopy.

INTRODUCTION

Loss of soybean production to pathogens is a serious problem in the United States. According to a report from North Central Regional Committee on soybean disease, the 3-year average loss due to pathogens (1989–1991) was more than 13% of total production accounting about 6.4×10^6 metric tons in the north central United States.^[1] Fungal damage occurs most frequently when warm and wet weather delays the soybean harvest.^[2,3] More than 40 bacteria, fungi, and viruses can infest soybean seeds. However, *Phomopsis*, *Personospora manshurca Syd* (downy mildew), soybean mosaic virus (SMV), *Cercospora kikuchii*, and *Fusarium graminearum* are the most economically important; and these fungal pathogens have a devastating effect on both chemical and physical properties of soybean; which limits end-use.^[4] Mbuvi et al. studied the effect of fungal damage on physical properties of soybean and found that fungal damage reduces seed density by 4%, volume and weight by 13%, and increases breakage susceptibility by 20 times compared with healthy seeds.^[5] Pathogens such as *Phomopsis* and *C. kikuchii* significantly affect oil quality. Research showed that oil produced from such infested soybean seeds has elevated free fatty acid concentration and low refining yield. The oil has darker color, lower oxidative stability, and poor flavors.^[6] Fungal-damaged soybean seeds may also have higher protein and lower carbohydrate contents compared with healthy soybean seeds.^[7,8] Fungal-damaged soybeans could contain some mycotoxins produced by some fungi. Aflatoxin, T-2 toxin, diacetoxyscirpenol (DAS), HT₂ toxin, low levels of vomitoxin (DON), zearalenone (F₂), and fumonisin (B₁) were also found in fungal-damaged soybean seeds.^[3] Fungal damage, therefore, is a critical factor for soybean grading and marketing.

Biological factors such as fungal damage, that affect soybean quality cannot be quantified easily and the criteria for these visual judgments have been defined only recently.^[9] Symptoms such as discoloration of a seed coat caused by fungal damage are usually used as an indicator of soybean quality and classification of the type of fungi. The maximum limits for damaged soybeans are $\geq 1, 2, 5,$ and 10% for US Nos. 1, 2, 3, and 4 designations, respectively.^[10] The maximum limit on fungal-damaged soybeans is very restricted, especially for US Nos. 1 and 2 soybeans. Fungal damage can be determined by physical properties of soybean seed coats using machine vision. Casady developed a color-image processing system to classify soybean seeds into healthy and symptomatic seeds infested by *Phomopsis*, *Alternaria*, and *C. kikuchii*.^[11] With this system, the fungal-damaged seeds were correctly determined in 77–91% of the soybean seeds tested. Ahmad et al. used a color classifier for classifying fungal-damaged soybean seeds.^[12] Their color analysis showed differences between the healthy and fungal-damaged seeds. The average classification accuracy was 88% for the healthy and fungal-damaged seeds. However, the classification accuracies for *Phomopsis*, *Alternaria*, *F. graminearum*, and *C. kikuchii* were only 45, 30, 62, and 83%, respectively.

Machine vision is the application of electronic imaging to enable the visual inspection of an object or a scene. 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. Wang et al. developed classification models to measure damaged soybean seeds using near-infrared (NIR) spectroscopy.^[13] The classification accuracies were 100, 98, 97, 64, 97 and 83% for sound seeds and those damaged by weather, frost, sprout, heat, and mold, respectively. The results indicated that NIR technology should be considered for the classification of fungal-damaged soybean seeds. The objective of this research was to classify healthy and fungal-damaged soybean seeds and discriminate among various types of fungal damage using NIR spectroscopy.

MATERIALS AND METHODS

Soybean samples were obtained from GIPSA Federal Grain Inspection Service (FGIS) (Kansas City, MO) and the Department of Agronomy, Kansas State University. Healthy and fungal-damaged seeds were manually classified by trained grain inspectors. Seeds of five categories were used for this study. The categories were healthy seeds (natural yellowish color), *Phomopsis* damaged seeds (white to gray mycelium and shriveled seed coat), *C. kikuchii* damaged seeds (violet to dark purple color and discolored areas ranged from scattered specks to entire seed surface), SMV damaged seeds (mottled with black or various shades of brown bleeding from the hilum), and downy mildew damaged seeds (superficial milky white encrustation of mycelium and oospores).

Single soybean seed spectra from 400 to 1700 nm were collected with a diode-array NIR spectrometer (DA7000, Perten Instruments, Springfield, IL). 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. Diode-array NIR spectrometer collected spectra at a rate of 30 s^{-1} . 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 500 healthy seeds and 800 fungal-damaged seeds infested by *Phomopsis*, *C. kikuchii*, SMV, or downy mildew (200 each) were measured, for a total of 1300 soybean seeds.

Partial least squares (PLS) (Galactic Industries, Salem, NH) and neural networks (NeuralWare, Inc., Pittsburgh, PA) were used for data analysis and model development. Soybean seeds first were separated equally into calibration and validation samples sets, based on even and odd numbers. Healthy and fungal-damaged seeds were assigned constant values of 1.0 and 2.0 for two-class models, respectively. A seed was considered to be correctly categorized if the predicted value lay on the same side of the midpoint of assigned values. For five-class models, healthy and fungal-damaged seeds infested by *Phomopsis*, *C. kikuchii*, SMV, or downy mildew were assigned constant values of 1.0, 2.0, 3.0, 4.0, and 5.0, respectively. For PLS models, 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. For neural network models, the determination of major parameters such as learning cycles, learning rate (η), and momentum (α), were described by Wang et al.^[13] The network with the highest validation accuracy was recorded as the best model. Visible (490–750 nm), NIR (750–1690 nm), and full wavelength (490–1690 nm) regions were used to develop classification models. Spectral data below 490 nm and above 1690 nm were discarded because of signal noise.

Grams/32 software (Galactic Industries, Salem, NH) was used for determination of soybean seed color. Reflectance spectra from 490 to 750 nm were transferred into $L^*a^*b^*$ color space using Grams/32 (Galactic Industries, Salem, NH) 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; b^* varies from -100 to 100 and is a measure of blue when negative and yellow when positive.

RESULTS AND DISCUSSION

The full wavelength region of 490–1690 nm gave the highest percentage of correct classification for both cross-validation and prediction ($>99\%$) (Table 1). The use of the visible or NIR wavelength region alone resulted in low classification accuracy. When compared to healthy soybean seeds, the energy absorption of fungal-damaged soybean seeds was higher in the visible region and lower in the NIR region (data not shown). The differences in energy absorption indicate differences in color and chemical composition between healthy and fungal-damaged soybean seeds.

Fungal-damaged soybean seeds are usually associated with discoloration and fissures on the seed coats. In addition, fungal damage also results in wrinkled, cracked, and crusty seed coats with less roundness. The degree of discoloration varies with type of fungi. The average spectra of healthy and fungal-damaged seeds showed that the energy absorption values of *Phomopsis* and downy mildew damaged seeds were lower than the healthy seeds in the visible wavelength region, where as *C. kikuchii* and SMV damaged seeds had higher energy absorption values than the healthy seeds in the NIR region (Fig. 1).

Table 1. Classification accuracies (%) of healthy and fungal-damaged soybean seeds using two-class PLS models.

Spectra region	F^c	Calibration results ^a			Validation results ^b		
		Healthy	Damaged	Average	Sound	Damaged	Average
490–750 nm	6	99.2	97.0	97.8	99.6	97.8	98.4
750–1690 nm	10	99.6	97.3	98.3	99.6	98.0	98.6
490–1690 nm	10	100	99.8	99.8	100	98.5	99.1

^aThe total number of soybean seeds in calibration sample set is 650.

^bThe total number of soybean seeds in validation sample set is 650.

^c F , number of PLS regression factors.

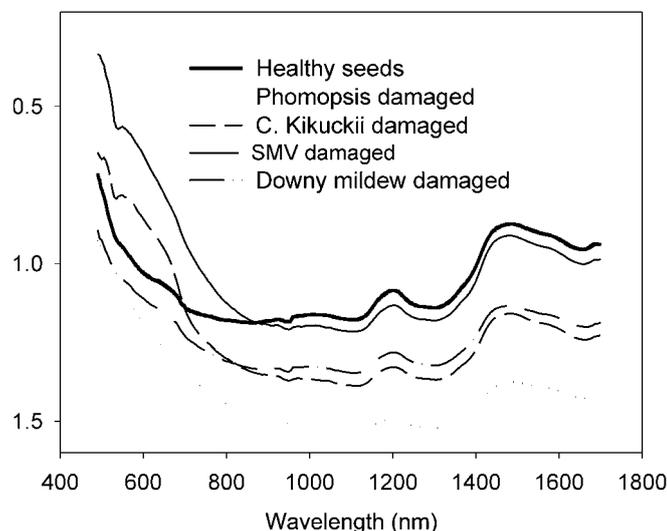


Figure 1. Average visible and near-infrared absorption curves for healthy soybean seeds, and *Phomopsis*, *C. kikuchii*, SMV, and downy mildew damaged soybean seeds.

Phomopsis and downy mildew damage result in white to gray and milky white color on the seed coat. *Cercospora kikuchii* and SMV damage results in violet to darker purple and black color on the seed coat. The color differences between the healthy and fungal-damaged seeds were further confirmed by *L*, *a*, and *b* values (Table 2). Healthy seeds had higher *L* values than *C. kikuchii* and SMV damaged seeds and lower *L* values than *Phomopsis* and downy mildew damaged seeds. This response indicates that *C. kikuchii* and SMV damaged seeds are darker than healthy seeds, and healthy seeds are darker than *Phomopsis* and downy mildew damaged seeds. Healthy seeds had higher *b* values than fungal-damaged seeds, indicating healthy seeds are more yellow than fungal-damaged

Table 2. The color variation between healthy and fungal-damaged soybean seeds measured as *L*, *a*, and *b* values in *L*a*b** color space.

Type of fungal-damage	<i>L</i>	<i>a</i>	<i>b</i>
Healthy seeds	37.27c ^a	7.59c	12.06a
Fungal-damaged seeds			
<i>Phomopsis</i>	44.84a	9.85a	11.85a
<i>C. kikuchii</i>	29.23d	7.13cd	7.73d
SMV	24.61e	8.43b	10.73b
Downy mildew	41.89b	6.56d	9.20c
Average	35.14	7.93	9.88
LSD	1.96	0.67	0.85

^aValues within the same column followed by different letters are significantly different at $P < 0.05$.

seeds. However, there were no significant differences between healthy seeds and SMV damaged seeds in a value.

The peaks and valleys of the beta coefficients curve from the PLS model illustrate the significant differences in energy absorption between the healthy and fungal-damaged soybean seeds (Fig. 2). The differences in energy absorption in the NIR region indicate differences in chemical composition between healthy and fungal-damaged soybean seeds. The beta coefficient peaks around 565, 650, and 705 nm are related to color differences between healthy and fungal-damaged seeds. The beta coefficient peaks around 970 and 1465 nm are related to NH bonds representing protein content.^[14] The peaks around 1330 and 1600 nm are related to CH bonds representing fiber and starch contents.^[15] The fungal-damaged soybean seeds may have higher protein and less carbohydrate contents than healthy soybean seeds.^[5,7,8,16] *Phomopsis* seed decay is the most destructive disease of soybean seeds.^[2,17,18] Severely affected seeds are elongated, heavily cracked or fissured and shriveled with less mass, volume, density, and more surface area compared to healthy seeds. *Cercospora kikuchii* damage reduces density or mass and causes degradation of seed-coat protein.^[19]

Preliminary results showed that the neural network without a hidden layer yielded higher classification accuracies for both calibration and validation sample sets.^[20,21] Thus, neural network models without a hidden layer were developed. The optimum neural network parameters, learning cycle of 150,000, learning rate of 0.7, and momentum of 0.5, were used for neural network training. The performance of neural network models is summarized in Table 3. In general, calibration and validation sample sets had similar classification accuracies. The neural network using the wavelength region of 490–1690 nm yielded the highest classification accuracy. More than 90% of healthy, *Phomopsis*, SMV, and downy mildew damaged soybean seeds were correctly classified for both calibration

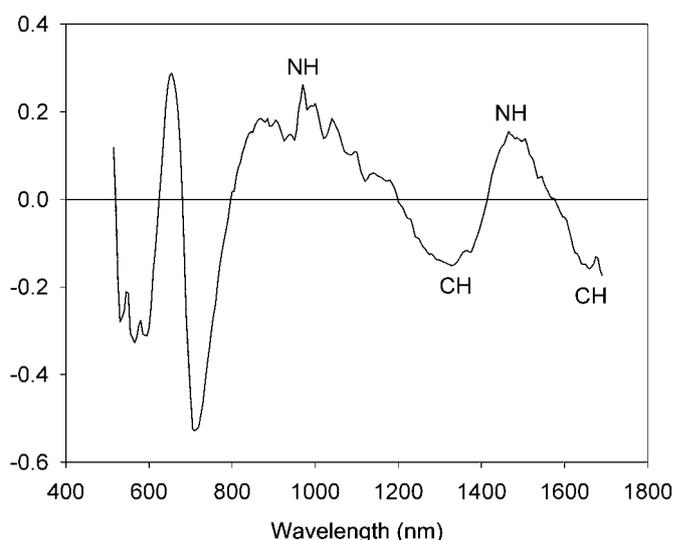


Figure 2. Beta coefficients curve of PLS model (10 PLS factors) for classification of healthy and fungal-damaged soybean seeds.

Table 3. Effect of wavelength region on classification accuracies (%) of healthy and fungal-damaged soybean seeds using five-class neural network models.^a

Sample sets	Classification accuracy (%)					Average (%)
	Healthy	<i>Phomopsis</i>	<i>C. kikuchii</i>	SMV	Downy mildew	
Calibration set ^b						
490–750 nm	99.6	98.0	80.0	90.0	88.0	91.1
490–1690 nm	99.6	100	83.0	95.0	90.0	93.5
750–1690 nm	99.2	100	54.0	95.0	67.0	83.0
Validation set ^c						
490–750 nm	100	96.0	76.0	91.0	91.0	90.8
490–1690 nm	100	99.0	84.0	94.0	96.0	94.6
750–1690 nm	100	100	53.0	94.0	71.0	83.6

^aParameters used: wavelength range: 490–1690 nm; momentum, 0.5; learning rate, 0.7; and learning cycle, 150,000.

^bThe total number of soybean seeds in calibration sample set is 650.

^cThe total number of soybean seeds in validation sample set is 650.

and validation sample sets, respectively. *Cercospora kikuchii* damaged seeds were the most difficult-to-classify. However, 83% and 84% of *C. kikuchii* damaged soybeans were correctly classified for calibration and validation sample sets, respectively, when the wavelength region of 490–1690 nm was used. About 16% of *C. kikuchii* damaged soybean seeds were misclassified into healthy and *Phomopsis* damaged soybean categories, and about 6% of SMV damaged seeds were misclassified into healthy seeds (data not shown).

CONCLUSION

Healthy and fungal-damaged soybean seeds can be easily classified using PLS models with NIR. Classification accuracy was more than 99% when the wavelength region of 490–1690 nm was used under a two-class model. Five-class neural network models were developed to discriminate among various types of fungal damage. The average of correct classifications was 93.5% for the calibration sample set and 94.6% for the validation sample set. The classification accuracies of the validation sample set were 100, 99, 84, 94, and 96% corresponding to healthy seeds, *Phomopsis*, *C. kikuchii*, SMV, and downy mildew damaged seeds, respectively.

REFERENCES

1. Doupnik, B.J.R. Soybean production and disease loss estimates for north central United States from 1989 to 1991. *Plant Dis.* **1993**, *77*, 1170–1171.
2. Sinclair, J.B.; Backman, P.A. *Compendium of Soybean Diseases*, 3rd Ed.; American Phytopathological Society, St. Paul, USA, 1989, pp. 38.

3. Wilson, R.F. Dealing with the problems of fungal damage in soybean and other oilseeds. *JAOCS* **1995**, *72*, 1413–1414.
4. Sinclair, J.B. Discoloration of soybean seeds—an indicator of quality. *Plant Dis.* **1992**, *76*, 1087–1091.
5. Mbuvi, S.W.; Litchfield, J.B.; Sinclair, J.B. Physical properties of soybean seeds damaged by fungi. *T. ASAE* **1989**, *32*, 2093–2096.
6. List, G.R. *Handbook of Soybean Oil Processing and Utilization*; American Oil Chemists Society: Champaign, USA, 1980.
7. Pathan, M.P.; Sinclair, J.B.; McClary, R.D. Effect of *Cercoospora kikuchii* on soybean seed germination and quality. *Plant Dis.* **1989**, *73*, 720–723.
8. Wilson, R.F.; Novitzky, W.P.; Fenner, G.P. Effect of fungal damage on seed composition and quality of soybeans. *JAOCS* **1995**, *72*, 1425–1429.
9. Sinclair, J.B. Reevaluation of grading standards and discounts for fungus-damaged soybean seeds. *JAOCS* **1995**, *72*, 1415–1419.
10. USDA. *Grain Inspection Handbook*; subpart J—U.S. standards for soybean. U.S. Department of Agriculture: Washington, D.C., USA, 1994.
11. Casady, W.W.; Paulsen, M.R.; Reid, J.F.; Sinclair, J.B. A trainable algorithm for inspection of soybean seed quality. *T. ASAE* **1992**, *35*, 2027–2033.
12. Ahmad, I.S.; Reid, J.F.; Paulsen, M.R. Color classifier for symptomatic soybean seeds using image processing. *Plant Dis.* **1999**, *83*, 320–327.
13. Wang, D.; Ram, M.S.; Dowell, F.E. Classification of damaged soybean seeds using near-infrared spectroscopy. *T. ASAE* **2002**, *45*, 1943–1948.
14. Shenk, J.S.; Workman, J.J.; Westerhaus, M.O. Application of NIR spectroscopy to agricultural products. In *Handbook of Near-Infrared Analysis*; Burnes, D.A., Ciurczak, E.W., Eds.; Marcel Dekker Inc.: New York, 2001; 419–474.
15. Barton, F.E., II; Burdick, D. Preliminary study on the analysis of forages with a filter-type near-infrared reflectance spectrometer. *J. Agric. Food Chem.* **1979**, *27*, 1248–1252.
16. Paulsen, M.R. Using machine vision to inspect oilseeds. *INFORM* **1990**, *1*, 50–55.
17. Sinclair, J.B. *Phomopsis* seed decay of soybeans—a prototype for studying seed disease. *Plant Dis.* **1993**, *77*, 329–334.
18. Velicheti, P.K.; Kollipara, K.P.; Sinclair, J.B.; Hymowitz, T. Selective degradation of protein by *Cercoospora kikuchii* and *Phomopsis longicolla*. *Plant Dis.* **1992**, *76*, 779.
19. Bochereau, L.; Bourguine, P.; Palagos, B. A method for prediction by combining data analysis and neural networks: application to prediction of apple quality using near-infrared spectra. *J. Agric. Engng. Res.* **1992**, *51*, 207–216.
20. Dowell, F.E. Neural network parameter effects on object classification and wavelength selection. *Optical Engng.* **1994**, *33*, 2728–2732.
21. Wang, D.; Dowell, F.E.; Lacey, R.E. Single wheat kernel color classification using neural networks. *T. ASAE* **1999**, *42*, 233–240.

Received December 3, 2002

Accepted March 3, 2003