# Classification of Hard Red Wheat by Feedforward Backpropagation Neural Networks

Y. R. CHEN,[1] S. R. DELWICHE,[1] and W. R. HRUSCHKA[1]

Cereal Chem. 72(3):317-319


## ABSTRACT

Because of breeding practices, visual determination of kernel morphology is becoming less dependable for classification of hard red wheat as winter or spring. Because of the price differential between these classes, it is important to develop rapid, accurate, and automatable alternative methods. This study was conducted to determine whether feedforward backpropagation neural networks applied to near-infrared (NIR) diffuse reflectance spectra of ground kernels could perform the classification. The NIR diffuse reflectance spectra (1,100–2,500 nm) of 2,403 ground hard red wheat samples representing the United States crop for 1987–1990 were used with commercially available neural network software. Mathematical pretreatments included $\log_{10}(1/\text{reflectance})$ and second differences of the log data. Networks with and without hidden layers were used with various subsets of the full spectral region as inputs. When developed on samples from the 1987–1989 crop years, the best neural network models yielded 97.0 and 96.8% accuracies for calibration and validation sets, respectively, utilizing the full wavelength range. This performance declined slightly to calibration and validation accuracies of 96.3 and 95.9%, respectively, when the wavelength range of 2,142–2,472 nm was used. When applied to the 1990 crop year, the prediction accuracies of the full and abbreviated wavelength range models were 95.1 and 95.6%, respectively. These models performed better than a previously reported principal component analysis with Mahalanobis distance classifier. Neural networks, combined with second difference pretreatment, should be a very useful component of a NIR-based classification system.


In the United States, classification of wheat has become increasingly difficult because of an increasing number of cultivars per class, more overlapping of growing regions of different classes, and more crossbreeding between cultivars belonging to two or more classes. Discrimination between hard red winter (HRW) and hard red spring (HRS) wheat is of particular importance, because of the volume of trade and the price differentials between these classes. The knowledge and experience required to accurately perform this classification is becoming too complicated for grain inspectors conducting visual inspection based on kernel morphology. Instrumentation that could classify wheat rapidly and with little training would be very useful to federal grain inspectors, traders, and millers.

We have been conducting research to develop techniques to differentiate these two classes by near-infrared (NIR) diffuse reflectance spectroscopy. The advantages of the NIR method are that it is rapid, does not require much sample preparation, and can be used in field measurements. Delwiche and Norris (1993) used NIR spectra of ground wheat to compare various discriminant analysis models. They calibrated the models with 1987–1989 crop samples and found that a five-factor principal component analysis with Mahalanobis Distance (PCA/MD) classifier was most accurate. The classification rate was 95% when the model was validated on the 1987–89 samples not included in the training set, and 92% when predicting a set of the 1990 samples.

Artificial neural networks are widely applied to pattern recognition problems. Examples are optical character recognition, image classification, target recognition, and speech recognition. Recently, they were found to be effective using spectral data to classify poultry for quality control on processing lines (Chen 1992, Park and Chen 1993) and to classify undamaged and damaged peanut kernels (Dowell 1994). Neural networks have some potential advantage over previously reported mathematical classification methods because they are able to discover and use nonlinear relationships without depending on the analytical expertise of the person developing the calibration model.

This study uses the samples and spectra used in Delwiche and Norris (1993) to assess the ability of neural networks to perform the winter-spring classification in non-mixed-class samples.

## MATERIALS AND METHODS

The samples and their spectra were those used by Delwiche and Norris (1993). The samples were from the 1987–1990 annual central United States hard red wheat crop surveys conducted by Doty Laboratories (Kansas City, MO). The calibration (training) set consisted of 50 samples from each of the two classes and from each of the 1987, 1988, and 1989 surveys (for a total of 150 HRW and 150 HRS samples). Each 50-sample group was representative of the larger set from which it was drawn, in terms of NIR-determined protein content, NIR-determined hardness, and state of origin. Testing was done on two sample sets: 1,325 remaining samples from the 1987–89 surveys (called the validation set), and 778 samples from the 1990 crop (called the prediction set). Diffuse reflectance spectra of ground samples were collected with a spectrophotometer (model 6250, NIRSystems, Silver Spring, MD) in a wavelength range of 1,100–2,498 nm at 2-nm increments and stored as $\log_{10}(1/\text{reflectance})$ before further processing.

Figure 1 outlines the architecture of a feedforward backpropagation neural network with one hidden layer. The circles represent processing elements (nodes) grouped into layers; the lines represent

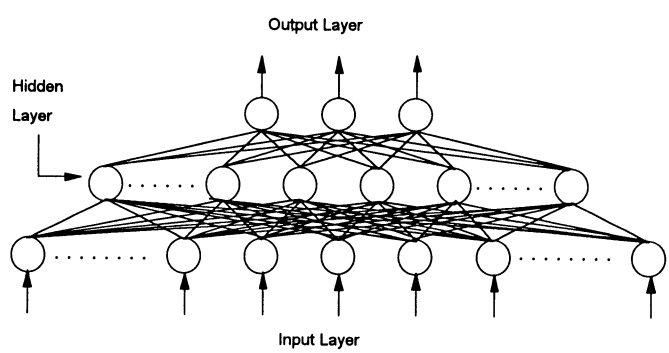


**Fig. 1.** Architecture of a feedforward backpropagation neural network.

[1]Research leader, agricultural engineer, and mathematician, respectively, Instrumentation and Sensing Laboratory, Beltsville Agricultural Research Center, ARS, USDA, Beltsville, MD.

Mention of a tradename or propietary product does not constitute a guarante or warranty by the USDA and does not imply its approval to the exclusion of other products that may be suitable.

the connecting weights. During the testing of a sample, the preprocessed spectral information is the output of the first layer, using one wavelength per node. The input to each node of the hidden and output layers is the weighted sum of the outputs of the nodes from the previous layer:

$$I_j = \Sigma w_{ji} o_i \quad (1)$$

where $I_j$ is the input to the $j$-node, $o_i$ is the output from the $i$-node, and $w_{ji}$ is the strength of the connection between the $i$-node and the $j$-node. A bias node with a constant input of 1.0 is added to each layer to allow for an offset of each weighted sum. The output of each node is determined from the input by the sigmoidal activation function:

$$o_j = 1/(1 + e^{-I_j}) \quad (2)$$

The spectral information is thus fed forward to produce the final classification information. During each training cycle (iteration), all samples in the training set are tested, and the differences (errors) between the outputs of the last layer and the desired outputs are calculated. The root mean square error (total error) of all the output nodes for all the training samples is then used to adjust all the weights in the network to make the total error smaller. Thus the total error is backpropagated through the network, using a gradient steepest-descent technique that iteratively minimizes the root mean square error between desired and actual network outputs. The rate of minimization is controlled by two constants: 1) learning rate, the fraction of the total error that is used to modify the weights; and 2) momentum, the fraction of the previous change in weights that is also used to modify the weights (helping to prevent overshooting a minimum total error). Both are dimensionless quantities. The initial set of weights is randomly generated. For detailed descriptions of the backpropagation mathematics, variants of the network described above, and other kinds of networks, see Pao (1989) or Hecht-Nielsen (1989). The network development software used was NeuralWorks Professional II/Plus (NeuralWare, Inc., Pittsburgh, PA), run in an MS-DOS environment.

Networks with and without one hidden layer were examined. The outputs of the first layer were either $\log_{10}(1/\text{reflectance})$ or their second difference values. These values were linearly normalized at each wavelength so that the minimum and maximum values of the training set values at that wavelength became 0 and 1, respectively. The second difference $S''(\lambda, g)$ of the spectrum at wavelength $\lambda$ was computed by:

$$S''(\lambda, g) = S(\lambda + g) - 2S(\lambda) + S(\lambda - g) \quad (3)$$

where $g$ is the gap measured in nanometers. The gap size was 20 nm, making the usable wavelength range 1,122–2,472 nm.

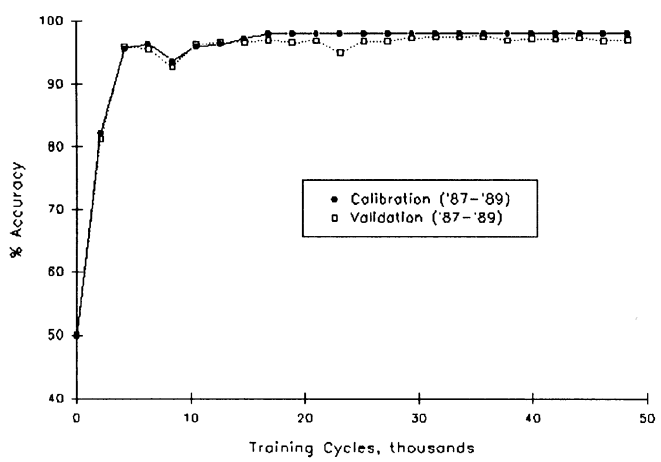


**Fig. 2.** Training and validation accuracies of the neural network with 226 input modes.

Because of memory size and computation time considerations, every third wavelength of either the $\log_{10}(1/\text{reflectance})$ or second difference spectra was used in the neural network, yielding a total of 226 first-layer nodes. The last layer had two nodes, one for each of the two decision classes, with the desired output values set to 0 for winter and 1 for spring.

The spectra in the validation set were used to monitor the performance of the training and to choose the optimal model. Initially, the accuracy of the neural net in classifying both the training and validation sets improved each time the training set was presented and the weights adjusted. After repeated training, the accuracy in classifying the validation set reached a maximum, while the accuracy in classifying the training set continued to increase. The best model was the one that classified the validation set most accurately. Beyond this point, the model was considered to be overtrained, and the neural network started to memorize the patterns in the training set rather than generalizing the patterns.

## RESULTS AND DISCUSSION

The networks with one hidden layer performed less well and converged much more slowly than the networks without hidden layers. Also, $\log_{10}(1/\text{reflectance})$ values without using second differences provided no significant separation of the spring and winter wheats. Therefore, only the results of second difference spectra with networks with no hidden layers are reported here.

Figure 2 is a plot of the training and validation accuracies of the neural network with 226 first layer nodes, using up to 50,000 training cycles. The learning rate and momentum were initially set to be 0.15 and 0.4, respectively. At 10,000 training cycles, they were reduced to 0.075 and 0.2, respectively, and they were further reduced to 0.0187 and 0.05, respectively, at 30,000 training cycles.

The accuracies (percentages of samples classified correctly) at the training cycles where the validation accuracy was highest are given in Table I. The accuracies for the training and validation sets were 98.0 and 97.6%, respectively, at 35,700 trainings. Table I also gives the results (prediction column) of applying the classifier to the 1990 crop (91.3%). The 1990 crop was not represented in the training set, which included 1987–1989 crops only. When the network was trained to predict the 1990 crop optimally, the best network was at 12,600 cycles, where the accuracy of the 1990 crop was 95.1%. At 12,600 training cycles, the average accuracies for calibration and validation sets were 96.3 and 96.6%, respectively.

The longer training time apparently provided a model that represented features more specific to those of the crop years used in training, thus the classification accuracy of the 1990 crop was reduced. For a network to also predict the crop not represented in the training set, the number of training cycles needs to be terminated when the validation set accuracy begins to approach the complement of the experimental error: 3% in the present study (Delwiche and Norris 1993).

Neural networks with no hidden layers using second differences with 113 first-layer nodes were also examined (Table II). Only the network with input range of 1,122–1,794 nm showed a slight reduction in the accuracies in calibration, validation, and prediction sets. In fact, during the training of this model, the calibration accuracy did not exceed 96.0%. All other models performed almost

**TABLE I**
**Confusion Matrix for Training Obtained from Feedforward Back Propagation Neural Networks with 226 Inputs and 2 Output Nodes (gap = 20 nm)**

| | Accuracy (%) | | |
|---|---|---|---|
| **Training Cycle** | **Calibration** | **Validation** | **Prediction** |
| 35,700 | 98.0 | 97.6 | 91.3 |
| 12,600 | 96.3 | 96.6 | 95.1 |

TABLE II
Comparing Different Networks with 113 Input Nodes

| Range of Wavelengths (nm) | Interval (nm) | Accuracy (%) | | | Cycles of Training |
|---|---|---|---|---|---|
| | | Calibration | Validation | Prediction | |
| 1,122–1,794 | 6 | 96.0 | 94.4 | 93.1 | 9,600 |
| 1,800–2,472 | 6 | 96.3 | 96.2 | 95.5 | 6,900 |
| 1,122–2,466 | 12 | 96.7 | 96.1 | 94.5 | 10,200 |
| 1,128–2,472 | 12 | 96.3 | 95.9 | 95.1 | 10,200 |

TABLE III
Comparing Different Networks with 56 Input Nodes with Input Wavelength Range of 330 nm

| Range of Wavelengths (nm) | Interval (nm) | Accuracy (%) | | | Cycles of Training |
|---|---|---|---|---|---|
| | | Calibration | Validation | Prediction | |
| 1,122–1,452 | 6 | 95.3 | 95.4 | 91.2 | 41,400 |
| 1,464–1,794 | 6 | 94.0 | 93.7 | 90.6 | 29,400 |
| 1,806–2,136 | 6 | 95.0 | 93.8 | 91.7 | 35,400 |
| 2,142–2,472 | 6 | 97.0 | 95.9 | 95.6 | 41,100 |

TABLE IV
Comparing Different Networks with 28 Input Nodes, Input Wavelength Range 162 nm

| Range of Wavelengths (nm) | Interval (nm) | Accuracy (%) | | | Cycles of Training |
|---|---|---|---|---|---|
| | | Calibration | Validation | Prediction | |
| 1,806–1,968 | 6 | 92.7 | 90.9 | 86.3 | 41,400 |
| 1,974–2,136 | 6 | 93.0 | 91.0 | 82.9 | 40,500 |
| 2,142–2,304 | 6 | 94.0 | 92.6 | 93.2 | 37,500 |
| 2,310–2,472 | 6 | 92.7 | 92.3 | 90.1 | 29,400 |

TABLE V
Results of Networks with Training 400 Ground Samples of 1987–1990 Crops, Validating on 2001 Samples of 1987–1990 Crops Not Included in the Training

| Range of Wavelengths (nm) | Interval (nm) | Accuracy (%) | |
|---|---|---|---|
| | | Calibration | Validation |
| 1,122–2,472 | 6 | 98.0 | 96.4 |
| 2,142–2,472 | 6 | 97.5 | 95.6 |

the same as when the whole spectrum was used (226 nodes, 1,122–2,472 nm, every 6 nm).

The number of input nodes was further reduced to 56, using various wavelength ranges of 330 nm (Table III). Validation accuracies were less than 96.0%, the best range being 2,142–2,472 nm with calibration and validation accuracies of 97.0 and 95.9%, respectively. The model also predicted the 1990 crop very well (95.6%). This implies that most of the spectral information needed for classification was at longer wavelengths. When the wavelength range and number of input nodes was further reduced to 162 nm and 28, none of the networks yielded validation accuracies above 92.6% or calibration accuracies over 94% (Table IV). Among the four ranges studied, the 2,142–2,304 nm yielded the best result.

To simulate the action of updating a neural network with a new crop year, 50 spring and 50 winter samples randomly selected from the 1990 crop were placed into the existing training set. Using the full wavelength range model, the network with 226 input nodes yielded an average calibration accuracy of 98.0% and validation accuracy of 96.4% (Table V). Validating on the 1990 crop alone (excluding the 100 moved samples), the accuracy was 95.5%. When the input wavelength range was limited to 2,142–2,472 nm, the accuracy was slightly reduced when compared to the results of the whole spectrum. However, predicting the 1990 crop alone, the accuracy of the abbreviated wavelength range model was slightly higher than the whole spectrum model (95.9 vs. 95.5%).

## CONCLUSIONS

Feedforward backpropagation neural network models without hidden layers are useful for classifying HRS and HRW wheat classes based on the NIR diffuse reflectance information of ground kernels. Calibrating on 1987–89 samples and validating on the same years' samples not used in the calibration yielded accuracies of ~97%, which approached the limit of the experimental data. The neural networks classifiers provided better classification of HRS and HRW cultivars than did the Principal Component Analysis with Mahalanobis Distance (PCA/MD) classifier previously reported.

Classification models based on 1987–89 samples could accurately (>95%) predict hard red spring and hard red winter cultivars of the 1990 crop, which were not included in the calibration. Updating the network with new crop year samples in the calibration set did not improve the prediction accuracy for predicting the new crop.

The spectral information in the long wavelength region of ground grain spectra was shown to be more useful for classification than that in the short wavelength region. The network based on the 2,142–2,472 nm wavelength region yielded accuracies that were nearly as high as networks based on the entire wavelength region (1,100–2,498 nm).

There remains the question of why the neural network model is giving better results than the models previously reported. Delwiche and Norris (1993) showed that NIR-determined protein (based primarily on protein absorbances) and NIR-determined hardness (based primarily on light-scattering information) used together could achieve accuracies of only ~85%. In contrast to these specific wavelength methods (which are fairly well understood as relating to well-defined absorbers or scattering effects), the partial least squares and discriminant analysis models (which use the full spectra and give higher accuracies) are more mathematically complex, yet less interpretable from a physical and chemical standpoint at this time. Although it is possible to see some of the protein and scattering effects in the intermediate spectra (loadings), nothing like a full analysis of the success of these popular methods exists in the literature. Neural networks are even less interpretable, because of the nonlinear transfer function, especially when hidden layers are used. Yet we think we have shown here that the neural networks are finding more information in the NIR spectra than simply protein and hardness. Examination of the neural network weight profiles and their relation to protein, starch, moisture, and lipids in wheat is currently under study.

## ACKNOWLEDGMENTS

We greatly appreciate the technical support of Minh T. Nguyen, Joyce D. Shaffer, and Ajay Dhuria.

## LITERATURE CITED

CHEN, Y. R. 1992. Classifying diseased poultry carcasses by visible and near-IR reflectance spectroscopy. Pages 46-55 in: Optics in Agriculture and Forestry. SPIE Vol. 1836. J. A. Deshazer, B. F. Fyodoroz, and G. E. Meyer, eds. Soc. Photo-Optical Instrumentation Eng.: Bellingham, WA.

DELWICHE, S. R., and NORRIS, K. H. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. Cereal Chem. 70:29-35.

DOWELL, F. E. 1994. Neural network classification of undamaged and damaged peanut kernels using spectral data. Optical Eng. 33:2728-2732.

HECHT-NIELSEN, R. 1989. Neurocomputing. Addison-Wesley: New York.

PARK, B., and CHEN, Y. R. 1994. Intensified multispectral image processing for poultry carcass inspection. Trans. ASAE 37:1983-1988.

PAO, Y.-H. 1989. Adaptive Pattern Recognition and Neural Networks, Addison-Wesley: New York.

[Received June 7, 1994. Accepted January 13, 1995.]