ULTRASONIC SIGNAL CLASSIFICATION FOR BEEF QUALITY GRADING THROUGH NEURAL NETWORKS †

A. Dale Whittaker *
Member ASAE

Bo Soon Park *
Member ASAE

James Darrell McCauley *
Student Member ASAE

Yanbo Huang *

Meat palatability is often related to the percent of intramuscular fat present in a meat cut. In fact, the USDA quality grades are based primarily upon "marbling", or the visual appraisal of intramuscular fat on a cut surface of a chilled carcass between the 12th and 13th ribs (USDA, 1980). Other factors influence palatability including animal age, amount of connective tissue, and aging effects; however, intramuscular fat is the primary component. This overall study was focussed on developing a robust method for measuring the amount of intramuscular fat in a beef cut non–invasively. This paper describes a portion of the research which compares statistical and neural network methods (supervised and unsupervised) for classifying the ultrasonic signal data.

A–mode, through–transmission ultrasound was used to non–invasively characterize the amount of fat in the lean tissue. Three ultrasonic frequencies were used including 1.0, 2.25, and 5 MHz. The radio frequency (RF) signals from the receiver were digitized for further analysis. These signals were then transformed into the frequency domain using a digital Fourier transform. Features from the frequency domain were extracted and used for the basis of the classifiers discussed in this paper. Previous research shows that the

---

†To be published in *Proceedings of the Automated Agriculture for the 21st Century Symposium*, Dec 16–17, 1991, American Society of Agricultural Engineers, Chicago, IL.

*Authors are in order Assistant Professor, Research Associate, and Graduate Research Assistants in the Department of Agricultural Engineering, Texas A&M University.

texture of the frequency domain power spectrum, or more specifically the number of local maxima, is strongly related to marbling in the meat sample (Park and Whittaker, 1990). As the percent of fat increases, the number of local maxima increases as well. The authors expect this to be due to the magnitude of random phase shifts based upon randomly spaced reflectors, and reverberations causing harmonic effects based upon consistently spaced reflectors.

The features of the power spectrum used for this paper are explained in detail in Park and Whittaker (1990) [1].

Future research will include the development of an input feature vector based more directly upon the shape of the power spectrum.

METHODOLOGY

Three independent approaches to developing a classifier were evaluated. These included a traditional statistical approach and two simulated neural network approaches (supervised and unsupervised learning). The classifiers were trained for all approaches using the same data set containing 100 individual meat samples. These samples covered the marbling ranges typically found in United States markets. In addition, the majority of the samples fell into the most important economic quality grades — USDA Select and Choice. The classifiers were evaluated using a separate data set of 24 samples. This validation data set had a very similar distribution to the training set.

Each approach was developed and verified by a different author. The data was shared, while the classification was done individually.

Statistical Approach

Multivariate linear regression models were developed from the training set and the most appropriate model was selected for validation. Comparisons were considered significant at the $P < 0.05$ level.

[1] The following abbreviations will be used in this paper to represent frequency parameters: FLO for low frequency, FHI for high frequency, FCUR for peak frequency, FCEN for central frequency, FHI for high frequency, FSKW for skewness, and NOLM for number of local maxima.

The basic problem for model selection is to find the simplest model possible which has a good fit. Since R^2 (multiple determinant coefficient) gives the proportion of the variability in the equation, a large multiple determinant coefficient is obviously desired. To achieve narrow confidence intervals and accurate estimates for these entities, the mean square error (MSE) should be small. The value of R^2 can always be increased by adding more terms to the model. However, the addition of unneeded variables may result in an increase in the mean square error. Thus, the real task is to balance these two measures of the goodness of the fit of the model.

The ideal model for prediction has a small PRESS (Prediction Sum of Square), small expected total error of estimation, small mean squared error (MSE), and large R^2. Stepwise regression selection was used to select the best model. Different criteria such as the coefficient of multiple determination (R^2), the total mean squared error (C_p), and the predicted sum of squares (PRESS) for comparing the regression models were used with the *all possible regressions* selection procedure.

Residual plots and partial regression plots were used for identifying influential outlying observations and multicollinearity. The studentized residuals, DFFITS values, and Cook's distance measurements were used to check for outlying and influential cases, and for model refinement and selection. Multicollinearity was studied by calculating the variance inflation factor (VIF). To identify outliers, the studentized residuals were examined for large absolute values.

Model validation usually involves checking the model against independent data. Since regression models are often used for a variety purposes, several different validation methods are commonly used. In this study, the cross–validation method was used to validate regression model.

A means of measuring the actual predictive capability of the selected regression model is to use this model to predict each case in the validation data set and then to calculate the mean of the squared prediction errors (MSPR).

$$\text{MSPR} = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)}{n} \tag{1}$$

where MSPR = mean squared prediction error,

Y_i = the value of the response variable in the ith validation case,

$\hat{Y}_i$ = prediction value for the ith validation case based on the model data set, and

n = the number of cases in the validation data set.

If MSPR is fairly close to the error mean square (MSE) based on the regression fit to the model data set, then MSE for the selected regression model is not seriously biased and gives an appropriate indication of the predictive ability of the model.

Supervised Neural Algorithms

Supervised learning algorithms use an external teacher or input/target vector (training) pairs. Usually, inputs are applied to the network and its output is calculated and compared to target values. The difference or error is fed back to the network and it adjusts its weights in a manner to reduce the error on the next iteration. The network is considered *trained* when the error is sufficiently low. Examples of supervised algorithms include popular backpropagation learning (Rumelhart et al., 1986) and Armstrong and Gesci's (1979) unconventional algorithm for adaptive logic networks.

There have been numerous successful applications of supervised learning for signal processing. A particularly relevant example using counterpropagation was presented by Roitblat et al. (1989). Using an artificial broadband pulse similar to a dolphin's click, transforming data from the time to frequency domain, and by binning frequency data, a neural network using amplitude information was 96.7% correct in distinguishing between three objects submersed under water. A dolphin, an animal with unique sonar capabilities, was only able to classify the three targets 94.5% correctly.

Despite many successes, supervised algorithms have been criticized as being biologically implausible (Hecht–Nielsen, 1989; Wasserman, 1989). In addition to this, algorithms can suffer from many pitfalls which make training difficult. Common pitfalls of backpropagation include network paralysis (input saturation caused by large weight values), local minima in the error surface, and improper choice of the learning constant or step size (Wasserman, 1989). Backpropagation and adaptive logic networks were the supervised algorithms chosen for this study. Each are briefly discussed below.

Backpropagation: Though a complete description of backpropagation is beyond the scope of this paper, a few descriptive remarks are necessary. The interested reader is referred to Wasserman (1989) for a general overview and to Hecht–Nielsen (1989) for a more mathematically rigorous treatment. The transfer function used was

$$O_i = \frac{1}{1 + e^{\left(\sum_{j=1}^{N} W_{ij} I_i - \Theta_i\right)}} \tag{2}$$

whose output is constrained between zero and one. Learning was governed by the momentum version of the generalized delta rule. The weight adjustment equation can be written as

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_j O_i + \alpha \left[\Delta W_{ij}(t)\right] \tag{3}$$

where α, the momentum coefficient, can be thought of as a "remembering" term. Tests of backpropagation were aided through the use of `neural.hpp`, a C++ library by Miller Freeman Publications[2].

Adaptive Logic Networks: Adaptive logic networks, which use only single logical functions AND, OR, and NOT, have certain advantages over backpropagation, especially in their hardware implementations. Dwelly (1990) briefly describes a software implementation of an adaptive logic network using binary trees with nodes of two types: adaptive elements and leaves. Adaptive elements can operate as an AND, OR, LEFT, or RIGHT logical gate, depending on its state. Leaf nodes serve only to collect inputs. Each tree produces a single boolean output. "For computing non–boolean outputs, several trees are used in parallel to produce a vector of bits representing the output value" (Dwelly, 1990). For details of this algorithm, see Armstrong and Gesci (1979).

Real valued estimates of marbling levels were obtained by dividing the interval of the real number line into 30 uniform quantization levels, essentially producing integer values. Most other parameters were divided into fewer levels. These integer quantization levels were encoded into 32–element vectors with a method designed to avoid producing functions which are sensitive to the least significant bit.

Investigations of adaptive logic networks were aided by `atree`, a C library by A. Dwelly and W.W. Armstrong of the Department of Computer Science,

[2] `neural.hpp`, © 1990, Miller Freeman Publications, was written by Larry O'Brien and discussed in the October 1990 issue of *AI Expert*.

University of Alberta[3].

Unsupervised Neural Algorithms

Unlike the backpropagation network whose training algorithm is a representation of supervised learning, the training algorithm of Kohonen's self–organizing network is a representation of unsupervised learning (Kohonen, 1984). In unsupervised training there is no "teacher." Input patterns are applied to the network and the network self–organizes by adjusting its weights according to a well–defined algorithm.

One important self–organizing principle of sensory pathways in the brain is that the position of neurons is orderly and often reflects some physical characteristics of the external stimulus being sensed. Kohonen presented such a self–organizing network which produces what he called self–organizing feature maps similar to those that occur in the brain (Lippman, 1987).

Kohonen's self–organizing network is a two–layer network. Its output nodes are orderly arranged in a one dimensional array in this case, and every input is connected to every output node by an adjustable connection weight. There are two primary reasons for such a network structure. One is that such a structure of the network is consistent with the problem we want to solve. The other is that it is easy to combine other networks with such a structure of the network.

The Kohonen self–organizing network has been successfully applied in areas such as speech recognition (Aleksander, 1989) and image processing (Nasrabadi and Feng, 1988).

There are two stages in the training process of Kohonen self–organizing network. The first stage is to generate a coarse mapping. In this stage the neighbors NE_{j*} are defined for updating as

$$NE_{j*} = (\max(1, j^* - 1), j^*, \min(m, j^* + 1)) \tag{4}$$

This means that every output node i has the neighbors $i-1$ and $i+1$, except at the borders of the array, whereby the neighbor of node 1 is 2, and the neighbor of node m is $m-1$, respectively. Also in this stage, in order to allow for large weight modifications and to settle into an approximate

[3] `atree`, © 1990, A. Dwelly and W.W. Armstrong, is available on INTERnet at `menaik.cs.ualberta.ca` (129.128.4.241).

mapping as quickly as possible, the adaptation gain term $\eta(t)$ is kept high (> 0.5).

The second stage is to fine–tune the output nodes within localized regions of the map to the input vectors. In this stage, the neighbors are reduced to winning nodes only, i.e. $NE_{j*} = (j^*)$, and the adaptation gain term, $\eta(t)$, should be low ($< 0.5$to $\ll 0.5$). In the training process $\eta(t)$ can be given in the form of

$$\eta(t) = Ce^{-t/T} \tag{5}$$

where C is a constant of maximum amount of change and should be close to 1, and T is a constant of decreasing rate of $\eta(t)$ whose value depends on the number of output nodes. To shorten the training process, both the input and weight vectors are normalized.

The algorithm was coded using the C programming language for common use as far as possible. When the program is run the first time, it is assumed that the training is in the coarse mapping stage. When the program is run the second time, it is assumed that the training enters the fine–tuning stage. The program can be run the nth ($n > 2$) time to fine–tune the training. So in every run, the program records necessary parameters such as the last weights, the number of training cycles of the running, etc., for the next running. In this way the program can be used flexibly, as many times as necessary to get acceptable training results.

After training, the program performs classification through applying an arbitrary input vector, calculating the excitation produced for each output node and then selecting the node with the highest excitation as the indicator of the correct classification.

RESULTS

The results are presented for each approach separately. Note that the statistical approach was actually a prediction equation, not a classification scheme. The prediction accuracy is evaluated using traditional approaches; however, for comparison with the neural network based classification approaches, the predicted values are clustered into classes. Using backpropagation, the values were first predicted using a single output node with little success. The results presented below are the result of classifying the output using three output nodes. The adaptive logic approach was also a predictive approach with the output being quantized predictions. The Kohonen approach is, perhaps, the most difficult to compare to the others. It is designed to group the data into "orthogonal" clusters. This does not guarantee that the clusters correspond to the desired divisions. For example, the output may cluster according to an important input variable that is not related to the amount of intramuscular fat (recall that there is no feedback mechanism in unsupervised training). Therefore, the classification boundaries that were selected may not be optimal for the Kohonen approach, but were imposed for the sake of comparison.

Several experiments were conducted for each approach to determine optimal parameter settings and network structures; however for comparison, each approach was evaluated according to a three level classification. The classes included the following intramuscular fat percentage ranges: $< 3\%$, 3–7%, and $> 7\%$. These boundaries were established according to important economic criteria (USDA grade divisions).

Statistical Approach

Through the analysis of frequency power spectrum data obtained by six different ultrasonic probes, it was shown that the determinant coefficients (R^2) for the shear probes were greater than those for the longitudinal probes. The highest determinant coefficient was recorded for the 2.25 MHz shear probe.

Also, note that the local maxima appears as the dominant term in all prediction equations. The multiple determinant coefficients of the model for predicting intramuscular fat using shear probes were larger than those for longitudinal probes. As mentioned in the methodology, the total mean squared error (C_p) was used for selecting the best model.

A multilinear regression model in the frequency domain was selected for each probe. The multiple determinant coefficients and their corresponding ranked variables for all ultrasonic probes are shown in Table 1.

In order to validate the multilinear model, the predicted intramuscular fat was calculated through the multilinear model, then those values were compared to actual ether extractable fat. From the validation data set, MSPR was 2.4573. Since the mean squared error (MSE) of the validation data was 2.5642 and the difference between these two values was only 0.1069,

Table 1: Determinant coefficient and corresponding ranked variables for selecting regression model to predict fat concentration in the frequency domain.

Probe (MHz)	Number in Model	R^2	C_p	Variables in Model
Longitudinal				
1	4	0.37	2.423	FHI, BNDW, FSKW, NOLM
2.25	3	0.41	0.786	FLO, BNDW, NOLM
5	3	0.55	0.503	NOLM, FHI, FCEN
Shear				
1	3	0.50	2.781	NOLM, FSKW, BNDW
2.25	3	0.82	1.595	NOLM, FLO, BNDW
5	2	0.58	1.128	NOLM, FCEN

the MSE for the selected multilinear regression model was not seriously biased and gave an appropriate indication of the predictive ability of the model.

In addition, paired T–tests were applied to test the hypothesis that no difference exists between actual fat concentration and predicted fat concentration from the model. As a result of the paired T–test, the mean difference was 0.6111 and the mean standard error was 0.3058 ($P < 0.05$) for the multilinear model in the frequency domain.

To show the difference between actual ether extractable fat and predicted fat numerically, error mean and standard deviation were calculated. The error mean of the model in the frequency domain was 1.1711 and the standard deviation was 1.0972. Therefore, it may be concluded that the multilinear model in the frequency domain was capable of predicting intramuscular fat with approximately 1.2 percent fat error.

Finally, according to the one–on–one based relationship plots between actual ether extractable fat and predicted fat, the multilinear model in the frequency domain seemed to be best for predicting low intramuscular fat (below 4 percent).

Supervised Neural Algorithms

Only the 2.25 MHz probe data was used for supervised training investigations. Convergence of the backpropagation network was not obtained with all seven frequency parameters so *a priori* information (statistical analysis) was used to select fewer input parameters.

Table 2 presents a summary of the results obtained with backpropagation. These outputs were processed by a *winner–take–all* network, i.e., the node with the largest output was considered the winner. The accuracy was calculated simply as the number of correct predictions divided by the total number of samples in the validation set.

Table 2: Classification of Marbling Levels with Backpropagation.

Input[a] Parameters	Training Pairs	Validation Samples	Classification Accuracy	CPU[b] (min)
Shear	9	24	54.2%	214
Longitudinal	9	24	70.8%	9.3
Shear	9	61	41.0%	150
Longitudinal	9	58	41.4%	6.1
Longitudinal	9	100	57.0%	28.5

[a] Parameters used were FCEN, FSKW, AND NOLM.
[b] For a Sun 4/490.

The results showed a more accurate classification of the first range (0–3% fat) with sharply declining accuracy as marbling levels increased.

Classification results for the adaptive logic network are given in Table 3. Increasing the quantization level increased the accuracy of encoding, but also slightly increased the mean error. Lower levels of quantization also slightly increased mean error.

Table 3: Classification of Marbling Levels with Adaptive Logic Networks

Input[a] Parameters	Classification Accuracy < 3	3–7	> 7	Overall	Mean[b] Error	CPU[c] (min)
Long: All Freq. Parameters	50.0%	61.5%	57.1%	58.3%	1.93	11.9
Shear: All Freq. Parameters	100%	53.8%	57.1%	62.5%	1.81	20.0
Long: FCEN, FSKW, NOLM	75.0%	61.5%	42.0%	58.3%	2.05	157
Shear: FCEN, FSKW, NOLM	100%	61.5%	57.1%	58.3%	3.85	5.1

[a]The number of training pairs varied from 93–97. Twenty four samples were used for classification.
[b]mean error $= \frac{1}{N}\sum(\text{actual} - \text{predicted})$ percent fat.
[c]For a Sun Sparcstation 1

Unsupervised Neural Algorithms

To evaluate the optimal effectiveness of the network, experiments can be performed according to the number of inputs, n (2–7 frequency parameters) and the number of output nodes, m (2–8 classes). There are 119 ($C_7^2 + C_7^3 + C_7^4 + C_7^5 + C_7^6 + C_7^7$) possibilities of n. So, there exist 833 (7×119) experiments. Obviously, to do so many experiments is not worthwhile; therefore, through experimentation and *a priori* statistical knowledge, important combinations were selected.

In the experiment, different combinations of 7 ultrasonic A–mode signal features were used and classed into 3, 4 and 8 classes as shown in Table 4.

According to Kohonen's recommendation, for good statistical accuracy the number of training cycles should be at least 500 times the number of output nodes. So, for different number of classes, i.e. number of output nodes the required numbers of training cycles are different. For instance, for 3 output nodes, the required number of training cycles should be 1500 (3×500); for 4 output nodes, the number should be 2000 (4×500); for 8 output nodes, the number should be 4000 (8×500). In the training the adaptation gain term $\eta(t)$ was taken in the expression of

$$\eta(t) = 0.99e^{-t/T} \tag{6}$$

Table 4: The Accuracies of Classification of Beef Ultrasonic Signal Data

No. of Input Parameters	3 Classes	4 Classes	8 Classes
7[a]	68.9%	63.5%	31.1%
6[b]	67.6%	54.1%	29.7%
5[c]	67.6%	54.1%	29.7%
4[d]	66.2%	54.1%	29.7%
3[e]	67.9%	58.1%	29.7%
2[f]	40.5%	23.0%	16.2%

[a] FLO, FHI, FCUR, FCEN, BNDW, FSKW, NOLM
[b] FHI, FCUR, FCEN, BNDW, FSKW, NOLM
[c] FCUR, FCEN, BNDW, FSKW, NOLM
[d] FCEN, BNDW, FSKW, NOLM
[e] BNDW, FSKW, NOLM
[f] FSKW, NOLM

For t=0, $\eta(0) = 0.99$, very close to 1. In order to divide the training process in 2 stages (coarse mapping, fine–tuning) reasonably for different required number of training cycles T should be different. For instance, for 3 output nodes, T should be 200; for 4 output nodes, T should be 250; for 8 output nodes, T should be 2000.

Comparison of Classifying Methods

Results have been presented on independent experiments using statistical and both unsupervised and supervised neural network approaches to classifying beef marbling levels according to frequency and time parameters. To compare each of these methods, the following benchmark test was performed:

> Using the 97 samples of all 2.25 MHz shear probe frequency parameters for training data, 24 samples extracted from the same data set before training were classified. Accuracy was determined for samples in the following ranges: $< 3\%$ fat, 3–7% fat, or $> 7\%$ fat. The number of *near misses* were recorded, where a near miss was defined as misclassification by $\pm < 0.5\%$ fat.

Table 5 shows the results of these benchmark studies.

Table 5: Benchmark Comparison of Statistical, Supervised, and Unsupervised Approaches

Actual Classes	Predicted Classes				Type II Accuracy
	< 3%	3-7%	> 7%	NM[a]	
< 3%	4	0	0	0	100.0%
3-7%	4	8	1	1	61.5%
> 7%	0	4	3	3	42.8%
Type I Accuracy	50.0%	66.7%	75.0%	–	63.9%

(a) Statistical Methods

Actual Classes	Predicted Classes				Type II Accuracy
	< 3%	3-7%	> 7%	NM	
< 3%	3	1	0	0	75.0%
3-7%	2	8	3	1	61.5%
> 7%	1	4	2	0	28.6%
Type I Accuracy	50.0%	61.5%	40.0%	–	55.5%

(b) Unsupervised Approach (Kohonen)

Actual Classes	Predicted Classes				Type II Accuracy
	< 3%	3-7%	> 7%	NM	
< 3%	4	0	0	0	100.0%
3-7%	6	6	1	2	46.1%
> 7%	2	2	3	0	42.9%
Type I Accuracy	33.0%	75.0%	75.0%	–	54.2%

(c) Supervised Approach (Adaptive Logic Network)

[a] NM is the number of *near misses* and is defined as a misclassification by less than ±0.5% fat.

DISCUSSION

The statistical approach performed best for the particular benchmark study selected. It was clear through experimentation that the neural network approaches performed better with different data sets and under different constraints. This indicates that the data set chosen apparently had few nonlinearities. The Type II accuracy (the most important from an economic standpoint) was high for the low fat cases and decreased for higher fat for all approaches. This is not entirely unexpected. The signal to noise ratio decreases as the fat in a sample increases. The authors believe this to be primarily attributable to Raleigh scattering. Close scrutiny of the data reveals that several of the training and validation pairs had fat values which fell very close to cutoff values. This may have influenced the misclassification and is somewhat reflected by the number of near misses in Table 5. Another source of error may have been the presence of outliers in the validation data set. Statistical outliers were removed from the training set (as discussed earlier) before any training occurred. Finally, the small data set sizes may have had too much variation for effective training to occur.

Besides extremely long training times, many pitfalls were experienced in attempting to train a backpropagation network. Because of apparent inseparability of input parameters, convergence was only obtained when using a subset of the data (three input parameters) and classifying them into three fat categories. The fact that all statistical outliers were not removed from the testing data may have decreased the overall accuracy rating; removing outliers from the training data may have reduced the networks capability to generalize. It is believed that data taken by an ultrasonic probe designed especially for beef tissue would produce a much better data set which would reduce training time of supervised training algorithms and also improve the performance of the trained network. The accuracy rating of 70.8% was encouraging, but when the trained network was supplied with more data, it was clear that the higher accuracy was the result of a fortuitous (though random) choice of validation data. Further investigation revealed that a majority of the correctly classified samples in the 70.8% trial had actual percentage fat values between zero and three.

Because of the lack of success with backpropagation, an alternative supervised training algorithm was sought. Subsequently, adaptive logic networks were investigated. Training times of adaptive logic networks were much shorter than that of backpropagation. Convergence was obtained either after a specified number of correct classifications in the training set was reached or after a limited number of iterations, which ever came first. The classification accuracy of adaptive logic networks were comparable or better than that of backpropagation.

The experimental results for unsupervised training show that more input variables result in more accurate classification. For a given number of output nodes, the accuracy of classification with all 7 parameters was the best, the accuracy with 2 parameters was the worst and the accuracies with 6, 5, 4 and 3 parameters were between these two extreme cases. This indicates that more input variables result in a reduced variability caused by the inputs. The 7–parameter input case can lower the noise level of the data and explain the output adequately, although there may be redundancy of input factors. Six, five, four and three parameter inputs can lower the data noise level and explain the output in varying degrees, but the 2–parameter input case cannot do these things acceptably. Because the 3–parameter input case is the turning point of the accuracy, which should result in the minimum redundancy of input factors, it is worth special attention.

Classifying beef ultrasonic signal data using the Kohonen self–organizing network was successful to some extent. However, as shown in other research, the network can only make the first approximate classification. To get more precise classification, other supervised methods should be incorporated, such as Kohonen's learning vector quantization (LVQ) (Kohonen, 1984), a Grossberg network (forming counterpropagation network), a backpropagation network (Wasserman, 1989), or a Hamming network (Lippman, 1987).

In conclusion, the data supplied to all classification approaches was relatively difficult to work with. The authors believe that a much larger data set is needed with the complete frequency power spectrum. Much information was hidden by using the seven features selected. More information can be communicated about the power spectrum by binning and integrating the areas within each bin. In addition, it may be concluded that for relatively linear data, statistical approaches are appropriate. Overall, the classification accuracies of the three approaches did not differ greatly. A combination of supervised and non–supervised neural network approaches should be pursued for this type of data.

REFERENCES

1. Aleksander, I. 1989. *Neural Computing Architectures.* The MIT Press.

2. Armstrong, W. W. and J. Gecsei. 1979. Adaptation algorithms for binary tree networks. *IEEE Transactions on Systems, Man and Cybertinetics* SMC-9:276–285.

3. Dwelly, A. 1990. An Implementation of Adaptive Logic Networks. Department of Computer Science, University of Alberta.

4. Hecht–Nielsen, R. 1989. Theory of the backpropagation neural network. *IEEE INNS International Joint Conference on Neural Networks* 1:593–605.

5. Kohonen, T. 1984. *Self–Organization and Associative Memory.* Springer–Verlag, Berlin.

6. Lippman, R. P. 1987. An introduction to computing with neural nets. *IEEE ASSP Magazine* pp. 4–22.

7. Nasrabadi, N. M. and Y. Feng. 1988. Vector quantization of images based upon the Kohonen self–organizing feature maps. *Proceedings of the IEEE International Conference on Neural Networks* 1:101–105.

8. Park, B. and A.D. Whittaker. 1991. *Ultrasonic frequency analysis for beef quality grading.* ASAE paper No 90–6573. The American Society of Agricultural Engineers, St. Joseph, MI 49085–9659.

9. Roitblat, H. L., Moore, P. W. B., Nachtigall, P. E., Penner, R. H., and Au, W. W. L. 1989. Dolphin echolocation: Identification of returning echoes using a counterpropagation network. *IEEE INNS International Joint Conference on Neural Networks* 1(1):295–300.

10. Rumelhart, D., G. Hinton, and R. Williams. 1986. Learning representations by backpropagating errors. *Nature* 323:533–536.

11. SAS. 1988. *SAS User's Guide: Statistics.* SAS Institute, Inc., Cary, NC.

12. USDA. 1980. *Official U.S. standards for grades of carcass beef. Food Safety and Quality Service.* Washington, D.C.

13. Wasserman, P. D. 1989. *Neural Computing: Theory and Practice.* Van Nostrand Reinhold, 115 Fifth Avenue, New York, NY 10003.