Breeding Science 48 : 257~ 262 (1998)

# Evaluation of Soybean Plant Shape by Multilayer Perceptron with Direct Image Input

**Mari Oide and Seishi Ninomiya**

*National Institute of Agro-Environmental Sciences*, Tsukuba, Ibaraki 305-8604, Japan
Present Address: *National Agriculture Research Center*, Tsukuba, Ibaraki 305-8666, Japan

**Summary**

**Evaluation of soybean plant shape in breeding is empirical and based on visual judgment, making it unstable and inefficient and pointing up the need for a quantitative alternative. Previous studies successfully applied evaluation by linear discriminant function, fuzzy logic or neural network, but these models required definition and selection of important features for judging shape. We developed a method based on a multilayer perceptron (MLP) with direct image input of binary soybean images which does not require any shape features. An MLP is a kind of neural networks, and can exhibit good performance in pattern recognition. A neural network is composed of units being simple processors, and connections between the units carrying numeric data from one unit to another. Units of an MLP are arranged on the layers, and connected each other between the adjoined layers. We used 326 soybean plant images judged either "Good", "Fair" or "Poor" by expert soybean breeders. The images were divided into supervisor and test data sets. We studied 175 different MLP structures, varying the number of layers, units and connections. After training each MLP with the supervisor data set, we evaluated matches between MLP output and breeder judgment with the test data set. The MLP with three layers, 8×8 input units, 16 hidden units and three output units proved to be the superior structure. Although performance in judgment was no higher than that of previous ones, our method has the decided advantage of not requiring definition and extraction of the shape features and may be applicable to other crops. We should note that the MLP structure is too complicated for us to understand the manner of breeders' empirical judgment through this model; that is, the MLP is almost a black box for us for the time being.**

**Key Words : image processing, multilayer perceptron, neural network, soybean plant shape, visual judgment.**

## Introduction

Visual judgment of plant shape plays an important role in breeding. Being primarily skill-and experience-based, however, such judgment tends to be somewhat subjective and unstable and efficiency depends strongly on the breeder. This points up the need for an automated alternative to introduce objectivity and stability.

Ninomiya and Shigemori (1991) defined several binary image features of soybean plant shapes for evaluating shape by image analysis, and suggested that a linear discriminant function based on these features could effectively replace visual judgment by experts in breeding. Ninomiya *et al.* (1992), Ambuel *et al.* (1997) and Oide *et al.* (1996) applied fuzzy logic and neural network models based on the same features of soybean shape and showed that these models also worked well. Because these models depend on shape features, however, we must redefine new features each time models are applied to a different crop. The feature definition process is inefficient.

In this study, we suggest a shape evaluation model based on the neural network of multilayer perceptrons (MLPs, Lippmann 1987) that accept the direct input of binary images instead of features parameterized from images. Several researchers have applied neural network models to biological and agricultural objects (Balfoort *et al.* 1992, Bordor *et al.* 1991, Brewster *et al.* 1992, Morris *et al.* 1992, Muskal and Kim 1992, Rataj and Schindler 1991, Stolorz *et al.* 1992, Vieth and Kolinski 1991, Zhang *et al.* 1992), but none enable direct image input. Our purpose was to develop an MLP whose judgments on soybean plant shape matched those given by breeders.

## Materials and Methods

We used the same binary images of 875 individual soybean plants as Ninomiya and Shigemori (1991), examining only 326 images for which three expert-breeders' judgments on shape agreed, in order to preserve data set objectivity. The judgments were classified as "Good", "Fair" or "Poor" (Fig. 1). The longest axis of each binary image crossing the plant silhouette was set as the vertical axis and image size was adjusted to 256×256 pixels. Each image was then scaled down to 8×8, 16×16, 32×32 and 64×64 pixels.

We applied MLPs with direct image input to judgment models (Fig. 2). An MLP is a kind of neural network models, and can exhibit good performance in pattern recognition. Neural networks are composed of simple processors (units) and communication channels (connections) between the units which carry numerical data from one unit to another. Units of an MLP are arranged on the layers, and only units on the adjoined layer are fully linked to each other and they never link over layers. These connections never form feedback loops. In

Received October 1, 1997. Accepted June 19, 1998.

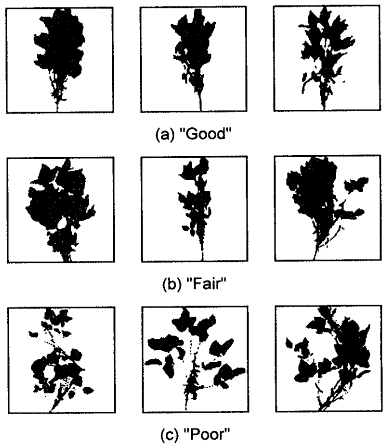


**Fig. 1.** Examples of binary soybean plant images.

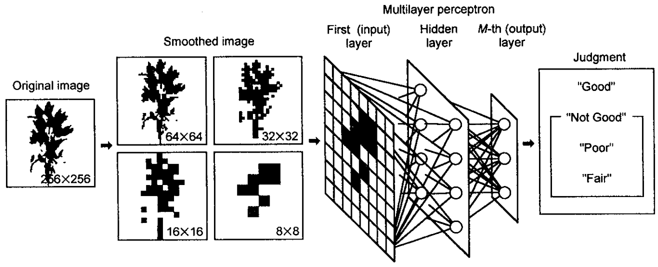


**Fig. 2.** Soybean plant shape evaluation by a multilayer perceptron. Binary images were directly input to the multilayer perceptron. $M$ indicates the number of the layers of the multilayer perceptron.

an MLP with $M$ layers, the first was an input layer. The layers between the second and $(M-1)$-th were hidden layers if they existed, and the $M$-th layer was an output layer. The output from the $j$-th unit of the $m$-th layer, $V_j^m$, is defined as follows:

$$V_j^m = 1 \qquad \text{for } j=0, \quad m=1,\ldots,M-1,$$
$$V_j^m = \xi_j \qquad \text{for } j=1,\ldots,n_m, m=1 \quad \text{and}$$
$$V_j^m = \sigma\left(\sum_{i=0}^{n_{m-1}} w_{ij}^m V_i^{m-1}\right) \text{ for } j=1,\ldots,n_m, \quad m=2,\ldots,M,$$

where $n_m$ is the number of the units of the $m$-th layer, $\xi_j$ is the intensity of the $j$-th pixel of an image, $w_{ij}^m$ is the connection weight from the $i$-th unit of the $(m-1)$-th layer to the $j$-th unit of the $m$-th layer, and $\sigma(.)$ is a sigmoid function defined as $\sigma(u)=\tanh u$. The output of each unit was set to a real number between $-1$ and $+1$. An MLP was a function composed of all the functions of the units, and the total number of the MLP parameters ($w_{ij}^m$'s) was

$$\sum_{m=2}^{M} n_m(n_{m-1}+1).$$

During the process of the MLP training, the MLP parameters ($w_{ij}^m$'s) were repeated to adjust so that error between the MLP outputs and the target outputs corresponding to the supervisor inputs became smaller. In this study, we set the supervisor inputs and the target outputs to be the soybean plant images and the breeders' judgments for them. The error of the $\lambda$-th supervisor data was evaluated by a function defined as:

$$E^\lambda(W) = \frac{1}{2}\sum_{j=1}^{n_M}\left(V_j^M(\xi^\lambda) - \zeta_j^\lambda\right)^2 \qquad \text{for } \lambda=1,\ldots,P,$$

where $P$ is the number of the supervisor data, $\xi^\lambda$ is the $\lambda$-th supervisor input vector, $\zeta_j^\lambda$ is the $j$-th component of the $\lambda$-th target output vector, and $W$ is the connection weight matrix. Note that the output of the each unit ($V_j^m$) was a function of $\xi^\lambda$. The MLP parameters ($w_{ij}^m$'s) were calculated by backpropagation (Rumelhart *et al.* 1986), which is one of the training rules for neural networks. In backpropagation, the parameters are given initial values, and then repetitiously updated based on the gradient descent rule, that gives

$$\Delta w_{ij}^m = -\eta\frac{\partial E^\lambda(W)}{\partial w_{ij}^m}$$

for $i=0,\ldots,n_{m-1}, j=1,\ldots, n_m, m=M,\ldots,2, \lambda=1,\ldots,P,$

where $\eta$ is the learning rate. In this study, connection weights were randomly initialized between $-0.3$ and $+0.3$ and the learning rate was set to be 0.2. Therefore, the backpropagation update rule was:

$$\Delta w_{ij}^m = \eta\,\delta_j^m(\xi^\lambda)\,V_i^{m-1}(\xi^\lambda) \text{ for } i=0,\ldots,n_{m-1}, j=1,\ldots,n_m, m=M,\ldots,2, \lambda=1,\ldots,P,$$

where

$$\delta_j^m(\xi^\lambda) = \begin{cases} \left(V_j^m(\xi^\lambda) - \zeta_j^\lambda\right)\sigma'\left(\sum_{i=0}^{n_{m-1}} w_{ij}^m V_i^{m-1}(\xi^\lambda)\right) \\ \text{for } m=M, \\ \left(\sum_{k=1}^{n_{m+1}} \delta_k^{m+1}(\xi^\lambda)\, w_{jk}^{m+1}\right)\sigma'\left(\sum_{i=0}^{n_{m-1}} w_{ij}^m V_i^{m-1}(\xi^\lambda)\right) \\ \text{for } m=M-1,\ldots,2, \end{cases}$$

in which $\sigma'(u) = \dfrac{d\sigma(u)}{du}$

We studied the effect of the number of layers, input units, hidden units and output units on successful identification rates (SIRs, matches between model output and breeder judgment). We also studied the effect of the number of connections between layers on SIRs. One hundred seventy-five MLP structures were studied (Table 1). The number of input units depended on the image size. As MLP input, the intensity of white pixels was to $-1$ and that of black pixels to $+1$. For three output units, target output pattern vectors for "Good", "Fair" and "Poor" were set to $(1, -1, -1)$, $(-1, 1, -1)$ and $(-1, -1, 1)$. For six output units, target output patterns were set to $(1, 1, -1, -1, -1, -1)$, $(-1, -1,$

**Table 1.** Network structures examined in this study; each cell shows the unmber of units on the layer

| layer ($M$) | input ($n_1$) | first-hidden ($n_2$) | second-hidden ($n_3$) | output ($n_M$) |
|---|---|---|---|---|
| 2 | either $8 \times 8$, $16 \times 16$, $32 \times 32$ or $64 \times 64$ | – | – | either 3 or 6 |
| 3 | either $8 \times 8$, $16 \times 16$, $32 \times 32$ or $64 \times 64$ | either 8, 16, 32 or 64 | – | either 3 or 6 |
| 4 | either $8 \times 8$, $16 \times 16$, $32 \times 32$ or $64 \times 64$ | either 8, 16, 32 or 64 | $n_2 \times$ either 1/2, 1 or 2 | either 3 or 6 |
| 3 | either $8 \times 8$, $16 \times 16$ or $32 \times 32$ | $n_1 \times$ either 1/2, 1 or 2 | – | square root of $n_1 \times$ either 3, 6 or 12 |
| 3 | either $8 \times 8$, $16 \times 16$ or $32 \times 32$ | either $8 \times 3$, $16 \times 3$, $32 \times 3$, or $64 \times 3$* | – | 3 |

* The unmber of connections between hidden and output layers was reduced. Hidden units were linked to a single output unit, and an output unit was linked from the same unmder of hidden units. In any other structure, hidden units were linked fully to output units.

1, 1, − 1, − 1) and (− 1, − 1, − 1, − 1, 1, 1). For more than six output units, target output patterns were set the same way. We defined a judgment by an MLP from the largest component of an output pattern. If an output pattern was either (0.1, 0.9, − 0.2) (three output units) or (0.1, 0.1, 0.9, − 0.8, − 0.2, 0.6) (six output units), for example, then the judgment by the MLP was defined as "Fair". We chose 26 images (7 "Good", 9 "Fair" and 10 "Poor") as a supervisor data set based on the typicality of the shape. Remaining data were used as a test data set.

The training of each MLP was repeated, changing initial connection weights randomly until 100 convergent cases were achieved. The number of trials to obtain 100 convergent cases depended on the network structure; some required only 100 trials, and others required more. These trails were terminated when no convergent case was attained for a week. This period was decided considering the computer performance (HP9000/750, Hewlett-Packard Co. Ltd., Palo Alto, CA). The convergence of each training was defined to be attained when the mean sum of the square of the deviations between the MLP output and the target output during the preceding 100 weight updates became less than 0.2. Unless the convergence was attained until 50,000 weight updates, the training was reset with new initial connection weights. We randomly sampled one from the supervisor data set for each weight update.

After 100 convergent cases, each MLP was evaluated by the test data set (300 images; 59 "Good", 84 "Fair" and 157 "Poor" images). In this study, a SIR denotes the rate that judgment by an MLP matches the judgment by breeders and was defined three following ways:

$$\mathrm{SIRA}=\frac{\text{(Number of successful identification of "Good" shape images)}}{\text{(Number of all images)}}\times 100$$

$$+\frac{\text{(Number of successful identification of "Not Good" shape images)}}{\text{(Number of all images)}}\times 100,$$

$$\mathrm{SIRG}=\frac{\text{(Number of successful identification of "Good" shape images)}}{\text{(Number of "Good" shape images)}}\times 100 \text{ and}$$

$$\mathrm{SIRNG}=\frac{\text{(Number of successful identification of "Not Good" shape images)}}{\text{(Number of "Not Good" shape images)}}\times 100,$$

where "Not Good" indicates either "Fair" or "Poor". We merged the "Fair" and "Poor" classes.

**Table 2.** Network structures (unmber of layers and units) not converged in training

| layer ($M$) | input ($n_1$) | first − hidden ($n_2$) | second − hidden ($n_3$) | output ($n_M$) |
|---|---|---|---|---|
| 2 | either 32 × 32 or 64 × 64 | − | − | either 3 or 6 |
| 3 | 8 × 8 | 128 | − | either 48 or 96 |
| 3 | 16 × 16 | 128 | − | 48 |
| 3 | 32 × 32 | either 512 or 2048 | − | either 96, 192 or 384 |
| 3 | 32 × 32 | 1024 | − | 384 |

## Results and Discussion

All network structures but those in Table 2 reached convergence in training. It took from 1 to 60 minutes to reach convergence on a workstation (HP9000/750, Hewlett Packard Co. Ltd., Palo Alto, CA). The number of weight updates until convergence varied between 400 and 1200 iterations, depending on the network structure. The more hidden units, the more rapidly training converged. Convergent efficiency accelerated as the number of input units increased. Whenever convergence was attained, learning was perfect; i.e., judgments at convergence in training were identical to those given by the breeders for the supervisor data set.

We examined the mean and standard deviation of SIRs of each network structure (Fig. 3-6). The variation in SIRs of each network structure was caused by different initial weight values. Notice that results for non-convergent cases cannot be plotted in figures. The means of SIRA, SIRG and SIRNG after removing outliers ((Quartile)±1.5×(Interquartile distance)) were distributed at 70-80%, 50-75% and 75-85%. The standard deviations of SIRA, SIRG and SIRNG were distributed at 2-10%, 7-27% and 4-18%. Smaller standard deviations indicated that SIRNGs were more stable than SIRGs. The means and standard deviations of SIRAs were close to SIRNGs, because "Not Good" images were much more dominant in the test data set (241 "Not Good" and 59 "Good" images).

We compared dependence on the number of layers, input units, output units and connections. For two layers (no hidden layers) with either 32×32 or 64×64 input units, no convergent case was attained (Table 2). All of three MLPs with four layers, 64×64 input units, eight first-hidden units and six output units showed extremely low SIR stability, even though convergence was attained in all such cases (Fig. 3-4). MLPs with 8×8 input units were more stable than all other cases. MLPs with more than about 100 times as many input units as hidden units (e.g., the case of the input layer with 64×64 units and the hidden layer with 32 units) tended to require more trials to obtain 100 convergent cases. They also showed low SIR stability. Increasing output units did not improve SIRs and made stability even worse (Fig. 5). No clear relationship was found between SIRs and the number of connections (Fig. 6). These results indicate that an MLP with too simple structure led underfitting, and convergence of learning was not good. On the other hand, an MLP with too complex structure led overfitting, and SIRs became low and unstable. An MLP with optimal structure must exist between them.

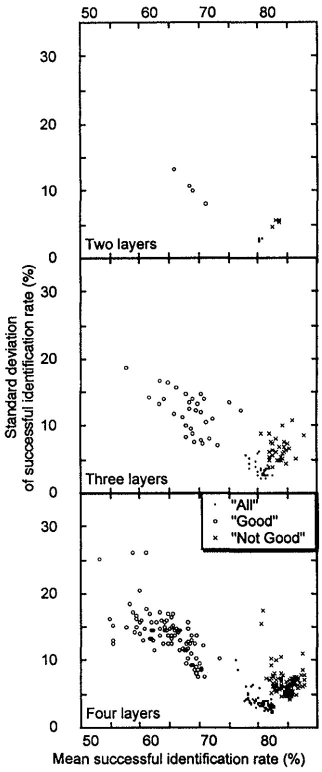


**Fig. 3.** Scatter diagrams of means and standard deviations of successful identification rates; the effect of the number of layers is compared.

The above results suggest that an MLP with three layers, 8×8 input units and three output units was

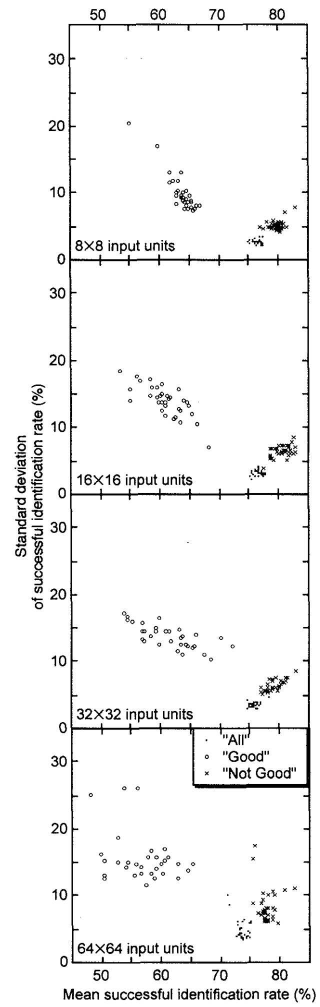


**Fig. 4.** Scatter diagrams of means and standard deviations of successful identification rates; the effect of the number of input units is compared.

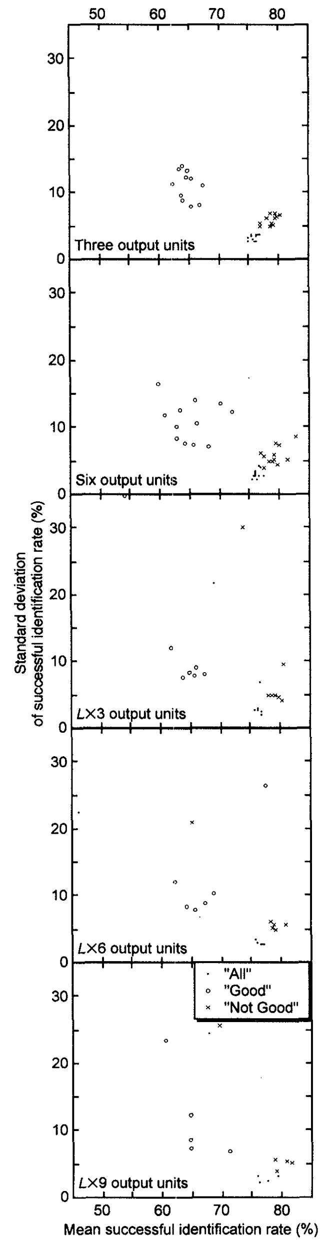


**Fig. 5.** Scatter diagrams of means and standard deviations of successful identification rates; the effect of the number of output units is compared. These diagrams were plotted only for multilayer perceptrons with three layers and either 8×8, 16×16 or 32×32 input units. *L* indicates the size of the side of an image; e.g., *L* is eight for an image with 8×8 pixels.

superior in training convergence and generalization ability. We found that the MLP with three layers, 8×8 input units, 16 hidden units and three output units showed the best performance because of the comparatively high averages and low standard deviations of SIRs (Fig. 7). The means of SIRA, SIRG and SIRNG were 76.1%, 63.8% and 79.2%, and the standard deviations were 2.76%, 8.90% and 5.16%, respectively.

We examined a direct image input model in this study. The method has the advantage that it does not require any shape feature extraction. The image, however, carries the whole information which may contain noises for the discrimination. Thus, its discriminant power is not supposed to be as high as that by the shape features being selected purely for the best discriminant performance. We, therefore, simplified the our problem, namely we decreased the number of the SIR classes into two classes considering the discrimination between "Good" and "Not good" shapes. We should evaluate also a SIR of "Fair" and "Poor" shapes in future. Our MLP structure was originally designed for three class problems and can be easily applied for such an expansion.

In this study, we found that MLPs with direct image input could attain quite good performance as substitutes for visual judgment on soybean shape. SIRs were no higher than those in previous studies (Ninomiya and Shigemori 1991, Ninomiya *et al.* 1992, Ambuel *et al.* 1997, Oide *et al.* 1996), but our proposed method has the decided advantage of not requiring extraction or selection of shape features. If we replaced visual judgment with previous models, we would have to define features for images and develop algorithms to extract

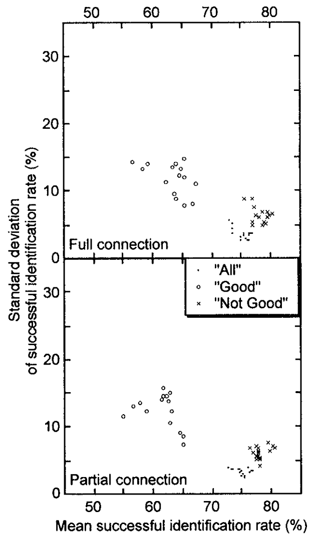


**Fig. 6.** Scatter diagrams of means and standard deviations of successful identification rates; the effect of the number of connections is compared. These diagrams were plotted for multilayer perceptrons with three layers, either 8×8, 16×16 or 32×32 input units and three output units.

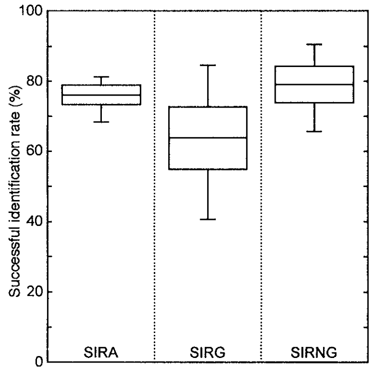


**Fig. 7.** Successful identification rates from the superior network structure (three layers, 8×8 input units, 16 hidden units and three output units). Boxes, horizontal lines in boxes and horizontal lines outside boxes show the range of mean ± standard deviation, the mean and the range.

them from images, and then select useful features for a target model. This extraction and selection must be repeated until the best model is found. The same process must be repeated to develop a shape evaluation model for different crops. Direct image input requires no extraction or selection. We expected the adaptability of this approach to other crops to be high.

We should note that the MLP structure is too complicated for us to understand the manner of breeders' empirical judgment through this model; that is, the MLP is almost a black box for us for the time being.

## Literature Cited

Ambuel, J., S. Ninomiya and N. Takahashi (1997) Fuzzy logic Evaluation of Soybean Plant Shape. Breed. Sci. 47: 253-257.

Balfoort, H. W., J. Snoek, J. R. M. Smits, L. W. Breedveld, J. W. Hofstraat and J. Ringelberg (1992) Automatic identification of algae: neural network analysis of flow cytometric data. J. Plankton Res. 14: 575-589.

Bordor, N., A. Harget and M. J. Huang (1991) Neural network studies 1. Estimation of the aqueous solubility of organic compounds. J. Am. Chem. Soc. 113: 9480-9483.

Brewster, M. E., M. J. Huang, A. Harget and N. Bordor (1992) Neural network studies 2. Use of a neural net to estimate oxidation energies for substituted dihydropyridines and related heterocycles. Tetrahedron 48: 3463-3472.

Lippmann, R. P. (1987) An introduction to computing with neural nets. IEEE ASSP Magazine 4: 4-23.

Morris, C. W., L. Boddy and R. Allman (1992) Identification of basidiomycete spores by neural network analysis of flow cytometry data. Mycol. Res. 96: 697-701.

Muskal, S. M. and S. H. Kim (1992) Predicting protein secondary structure content a tandem neural network approach. J. Mol. Biol. 225: 713-727.

Ninomiya, S., J. Ambuel, S. Morinaga and N. Takahashi (1992) Comparison of three models to evaluate soybean plant shape using shape indices given by image analysis: statistical discriminant function, neural network, and fuzzy logic. Proc. 7th Int. Congr. SABRAO and Int. Symp. WSAA: 613-620.

—— and I. Shigemori (1991) Quantitative evaluation of soybean (*Glycine max* L. MERRILL) plant shape by image analysis. Jpn. J. Breed 41: 485-497.

Oide, M., S. Morinaga and S. Ninomiya (1996) Selection method of supervisor data set to develop neural network evaluator on soybean plant shape. J. JASS 12 (1): 13-20, in Japanese with English summary.

Rataj, T. and J. Schindler (1991) Identification of bacteria by a multilayer neural network. Binary Comput. Microbiol. 3: 159-164.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986) Learning representations by back-propagating errors. Nature 323: 533-536.

Stolorz, P., A. Lapedes and Y. Xia (1992) Predicting protein secondary structure using neural net and statistical methods. J. Mol. Biol. 225: 363-377.

Vieth, M. and A. Kolinski (1991) Prediction of protein secondary structure by an enhanced neural network. ACTA Biochimica Polonica 38: 335-351.

Zhang, X., J. P. Mesirov and D. L. Waltz (1992) Hybrid system for protein secondary structure prediction. J. Mol. Biol. 225: 1049-1063.