Prediction of physical properties of orchid seedlings *'Phalaenopsis Sogo Vivien F819'* in a flask by digital imaging

C.-L. Hsieh[1]* and S.-F. Weng[2]

*[1]Department of Biosystems Engineering; and [2]Department of Machinery Engineering, National Pingtung University of Science and Technology, Pingtung, Taiwan, R.O.C. *Email: chinglu@mail.npust.edu.tw*

Hsieh, C.-L. and Weng, S.-F. 2005. **Prediction of physical properties of orchid seedlings *'Phalaenopsis Sogo Vivien F819'* in a flask by digital imaging**. Canadian Biosystems Engineering/Le génie des biosystèmes au Canada **47**: 3.23-3.32. Thanks to the improvement in cultivation technology and market expansion, the orchid seedling industry has become more and more important for Taiwan. In this research a non-destructive method, i.e. machine vision with a back lighting unit, has been used to acquire digital images of flask seedlings *Phalaenopsis Sogo Vivien F819* at stage 5. Sixteen image features extracted from each seedling were correlated to six physical properties (area, height, width, root length, fresh mass, dry mass) of flask seedling, to formulate property measurement models. A total of 297 samples were tested by five regression models and four back propagation neural network (BPN) models. Among them, 226 were used as the calibration set while the other 71 samples were used as the prediction set. The prediction results indicated that the multi-dimensional linear model (MDL), on average, had the highest coefficient of determination (R^2) among five regression models and showed an average prediction error of 0.401 for six properties. The back propagation neural network models with no hidden layer, one hidden layer with two or four nodes, and two hidden layers with 8-5 nodes were tested and the results showed that the two hidden layer model showed the highest prediction accuracy of 88.2% in area prediction and had 0.359 prediction error on average. Results also indicated that the BPN model performed about 10% better in prediction accuracy than the MDL model. These findings should be helpful for orchid farmers in production. **Keywords:** orchid, flask seedling, regression model, back-propagation neural network model, prediction, machine vision.

Grâce aux améliorations des techniques de culture et à l'expansion des marchés, l'industrie des plantules d'orchidées est devenue de plus en plus importante pour Taiwan. Dans cette étude, une méthode non-destructive, c.à.d. un système de vision artificiel avec une unité lumineuse en arrière plan, a été employée pour obtenir des images digitales de plantules en flacon de Phalaenopsis Sogo Vivien F819 au stade 5. Seize caractéristiques extraites des images de chacune des plantules ont été associées à six propriétés physiques (surface, hauteur, largeur, longueur de racine, poids frais, poids sec) des plantules en flacon pour élaborer des modèles de mesure des propriétés. Un total de 297 échantillons ont été testés par cinq modèles de régression et quatre modèles de réseau neuronal à rétropropagation (RNR). Parmi ceux-ci, 226 ont été utilisés pour réaliser la calibration des modèles et 71 échantillons ont servi à la vérification des modèles. Les résultats des modèles de prédiction ont indiqué qu'en moyenne le modèle linéaire multi-dimensionnel (LMD) présentait le coefficient de détermination (R^2) le plus élevé parmi les cinq modèles de régression et ont résulté en une erreur de prédiction moyenne de 0.401 pour les six propriétés. Les modèles de réseau neuronal à rétropropagation sans couche cachée, à une couche cachée avec deux ou quatre nœuds et à deux couches cachées avec 8-5 nœuds ont été testés et les résultats ont montré que le modèle à deux couches cachées fournissait la précision la plus grande (88,2%) pour la prédiction de la surface et présentait en moyenne une erreur de prédiction de 0,359. Les résultats ont également indiqué que le modèle RNR présentait une précision de 10% supérieure à celle du modèle LMD. Ces résultats devraient être utiles pour les producteurs d'orchidées. **Mots clés**: orchidée, plantule en flacon, modèle de régression, modèle de réseau neuronal à rétropropagation, prédiction, vision artificielle.

INTRODUCTION

Orchid (or *Orchidaceae*) has been one of the most popular house plants since the 19th century. It was not so popular before that because of the limitations of available cultivation technology. Today its popularity has grown rapidly, due to the improvement in propagation techniques and practices as well as the acceptance of orchid as a profitable crop by farmers and the efficient segmentation of supply chains (Griesbach 2000; Britt 2000). Thus, orchid has become one of the largest families of flowering plants in the market. Botanists who study these plants estimate that there are 17,000 to 35,000 species; according to well-documented reports there are around 20,000 species (Perry 2003). Statistical data of 1983 show that *phalaenopsis* (a popular variety of orchid) occupied less than 5% of the market in the Netherlands; while in 1994 its market share soared to over 66%. Large-scale potted phalaenopsis production is now taking place in China, Germany, Japan, the Netherlands, Taiwan, and the United States (Griesbach 2000). The world export/import trade of orchid plants and cut flowers exceeded CAN$198 million in year 2000, according to United Nations Comstats. Of this, about CAN$30 million was in orchid plants and about CAN$168 million in cut flowers. Thailand is the world's largest cut flower exporter while Taiwan is the largest orchid plant exporter (Laws 2004).

In Taiwan, orchid has been cultured for more than a century and has become an important export plant in recent years thanks to the significant improvement in cultivation technology and application of environmental control facilities. According to estimates, there are about 300 orchid farms in Taiwan and nearly 10% of them are related to production of orchid seedlings. The wholesale of the orchid industry in Taiwan reached about CAN$156 million in 2001 and is still increasing (Chen 2002). Orchid can be sold at different stages: seedlings

in flask (called flask seedlings), pot plant, blooming plant in pot, and cut flowers. To save space in transportation, follow trade regulations, and meet customer requirements, flask seedlings are a popular form for export in Taiwan. The quality of flask seedlings becomes an important issue for farmers because it will affect the commodity price and commercial reputation. To ensure the quality of seedlings, farmers need to inspect the flask seedlings one by one. This labor-intensive work is prone to manual error and needs to be improved to ensure competitiveness.

Usually, orchid can be propagated by seed or by tissue culture. These propagation procedures involve the following steps: 1) determine if a seedling has the qualities that make it worth being used as a parent, i.e. used as hybrid in a cross; 2) pollinate the flower; 3) let the seed capsule ripen and disperse seeds, followed by sowing seeds on agar in a flask; 4) when the seeds germinate and develop into protocorms, cut them and transplant them to new medium; 5) let the protocorms grow into seedlings; and 6) move the seedlings out from flasks and plant them into trays or pots, so they will grow into plantlets and commercial plants. In addition, steps 2) and 3) are analyzed using the seed propagation approach, while in the tissue culture approach these two steps would be replaced by cutting explants (usually the shoots of inflorescence) from the mother plant to be planted into agar in a flask and growing the explants into protocorm-like bodies (a pre-form of seedling). No matter which approach is applied in orchid propagation, the condition of the seedling in the flask is always one of the major concerns. It is always necessary to know the physical properties of the seedling, such as the mass, height, and root length for adjusting the nutrients and environmental parameters for the subsequent subculture procedures. Basically, it takes three or four subculture processes in an aseptic condition before the seedling can finally be transplanted to a small pot or tray. The final stage of subculture in a flask (called stage 5 herein) is an important stage because it is the final market product for flask seedlings. The price is decided by the quality of the product at this stage. Thus, farmers are deeply concerned about the quality of seedlings at stage 5. At present, the quality inspection of seedlings in stage 5 is a manual process and the accuracy depends on the inspector's experience. Picking out the seedlings and measuring their physical properties by instruments is an alternative, but this destructive approach limits its feasibility in practice. Therefore, using a non-destructive approach, for example machine vision, to measure the physical properties of seedlings attracts many researchers' attention.

Chen et al. (2002) indicated that machine vision has been applied in many agricultural sectors, such as land-based and area-based remote sensing for natural resource assessments, precision farming postharvest product quality and safety detection, classification and sorting, and process automation. Visen et al. (2004) applied image analysis to extract more than 150 features of five types of grains including barley, oats, rye, wheat, and durum wheat to identify the unknown grain types. Thomas et al. (2002) classified the northern peatland complex by using spectral image and plant community data. Yang et al. (2002) used digital images and a back-propagation neural network model to recognize weeds. Chen et al. (2001) measured the change in soybean plant cross-sectional area under wind conditions via image processing. Application of physical properties measurement of plants via digital images includes measurement of the canopy of tomato seedlings by machine vision (Ling and Ruzhitsky 1996). Jeng and Lin (1997) used machine vision to correlate the image features of a cabbage seedling with its physical properties, such as fresh mass of stem and leaves, total area of leaves, and dry mass of seedling. They also suggested a formulation to predict these properties by image features. Hsieh et al. (1997) applied image processing technique-texture analysis to cabbage seedlings grown in a tray. Five different growing stages could be classified by texture features and neural network classifiers, according to their study. Chen and Lin (1999) correlated the dry mass of orchid seedlings with leaf length and width. They also pointed out that there was no significant difference among species of orchid. Huang and Lin (2000) developed an algorithm to correlate the distance of leaves, angle between leaves, and length and width of leaves for orchid plantlets by image features. By using these image features, the orchid plantlet in a pot could be classified. Lin et al. (2001) developed a 3D graphical model to measure the growth curve of pepper seedlings. Although many studies have been conducted to explore the feasibility of prediction of plant seedling properties by machine vision and image processing techniques, research on physical property measurement of flask seedling is still very limited.

Phalaenopsis orchid is a promising product and reached the top in growth rate in year 2001, according to the auction report in the Netherlands (Chen 2004). It is also the most popular orchid produced in Taiwan. Therefore, this study has the following objectives:

1. To design a machine vision system for physical property measurement on flask seedlings of *Phalaenopsis* orchid.
2. To calibrate the image features and the physical properties of flask seedlings by a regression model and a back propagation neural network model.
3. To study the prediction accuracy of these models and to explore the reasons for their differences.

The results of this study could supply useful information for flask seedlings of *Phalaenopsis* orchid so as to assist farmers in quality control and culture improvement.

MATERIALS and METHODS

Sample preparation and image calibration

Samples of *Phalaenopsis, var. SOGO Vivien F819* were randomly collected from an orchid nursery farm located in the south of Taiwan. They were all at stage 5 and packed 25 seedlings per flask. After these samples were delivered to our laboratory, they were transplanted to a flask with new agar medium at one seedling per flask to avoid the image distortion caused by other seedlings. This one seedling per flask format sometimes has been conducted in the selection process for a new variety. Stage 5 is the last stage at which seedlings are a pot plant. Therefore, seedlings at this stage were the most mature compared to other stages of flask seedling.

A machine vision system for physical property measurement of a flask seedling (shown in Fig. 1) was developed. The system consisted of one CCD (2703N, SALKAI, Japan) with a lens (SPACECOM 6-12V, Tokyo, Japan), one lighting unit with two florescent lights (18W, TFC, Taiwan) inside, one personal computer of Pentium III CPU and 256 M RAM with a frame grabber (PCI-1411, National

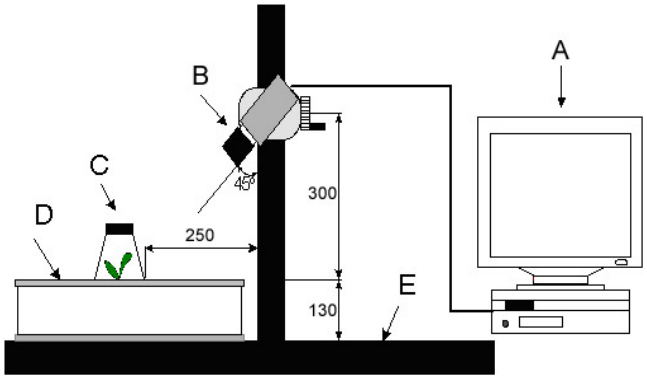

Fig. 1. Schematic diagram of image capturing system for orchid seedling in flask. A: personal computer; B: CCD camera; C: flask seedling; D: lighting unit; E: stands. Dimensions in millimeters.

Instruments, Austin, TX). The resolution of the image was 640×480 pixels of 24 bit in RGB. Before taking the pictures of flask seedlings, we took an image of a 2500 mm^2 square plate as the reference for further analysis. Each sample was imaged at two orthogonal directions, perpendicular to and parallel to the long axial direction of leaf extension. After the pictures were taken, we picked out the seedling and measured its six basic physical properties, i.e. area of plant (Y_1), height of plant (Y_2), width of plant (Y_3), root length (Y_4), fresh mass (Y_5), and dry mass (Y_6). The area of the plant was measured by an area meter (LI-COR Biosciences, Lincoln, NE) while the height, width, and root length of plants were measured by caliper (505-601, Mitutoyo, Kawasaki, Japan) with resolution 0.02 mm. Fresh mass and dry mass were recorded from the readings of an electronic balance (PG-203-S, METTER TOLEDO, Gresfinsee, Switzerland). Before the dry mass was measured, the fresh plant was dried by an ambient dryer (TTH-C3P, Ten-Billion,Taiwan) at 105°C for 24 hours (ASAE 2001).

General conditions of samples

A total of 297 samples were taken on seven sampling days from March 2002 to March 2003. From them, 226 samples taken from the first to the sixth sampling day were used as a calibration set for model setting while the remainder of them (71 samples) were used as the prediction set. Table 1 shows the six basic physical properties of flask seedlings. The average area of the seedlings was 1485 mm^2 with an 819 mm^2 SD. Properties were chosen based on the convenience and feasibility of measurement as well as their influence in market value. For instance, farmers favored a more healthy plant which is heavier and taller so as to have a better price. The standard deviation (SD) shows the variance of each property while the coefficient of variation (CV) indicates the variance in a standardized base. Results in Table 1 show that dry mass has the largest CV value of 0.67 followed by fresh mass and area of whole plant. Seedlings, when they were transplanted into the medium at stage 5, were not identical, although they were selected to be as uniform as possible. These differences among plants resulted in the variance of seedlings in physical properties. On the other hand, from beginning to the end of stage 5 will take two to three months, depending on the market requirements. The samples were selected randomly in stage 5. Thus the subculture time in the flask at this stage may also affect the properties of a seedling and result in this variance. Additionally, the cap of the flask also caused shading and affected the growth of plants.

Feature extraction

The image of a seedling taken by the machine vision system was in a 256 RGB color format and needed to be processed to a binary image before its features were extracted. These processes included gray-level image transform, binary image transform, and noise removal (shown in Fig. 2). The sample was asymmetric in shape and its leaves were grown in a long axial direction; from this direction, the width and height of the plant could be observed. Meanwhile, in the short axial direction, the width and overlapping condition of the leaf could be measured. Thus, each seedling was imaged from these two orthogonal directions (long axial and short axial directions) yielding a front-view image (FVI) and a side-view image (SVI). Eight features (X_1-X_8) were then extracted from these two images. The features were area, height, width, and circumference of FVI and of SVI, respectively. Another four features (X_9-X_{12}) were obtained by taking the average of FVI and SVI in area, height, width, and circumference, respectively. The area product of FVI and SVI, width product of FVI and SVI, height product of FVI and SVI, and square root of area product of FVI and SVI were also calculated as the other four features (X_{13}-X_{16}), giving a total of 16 features. Each feature was taken as the ratio to the 2500 mm^2 calibration plate in corresponding image. For instance, when the area of calibration plate in FVI was 200,000 pixels while the area of seedling was 240,000 pixels, the X_1 (area of FVI) feature was set to 1.20 (240000/200000=1.20). The image processing and feature extraction were conducted by IMAQ, a commercial image processing software (IMAQ 1999).

Model regression

Regression is one of the methods for studying the relations between dependent and independent variables. Five regression models were tested in this study. They were one-dimensional linear model (ODL model, Eq. 1), one-dimensional second-order polynomial model (ODP model, Eq. 2), multi-dimensional linear model (MDL model, Eq. 3), exponential model (EXP model, Eq. 4), and natural log model (LN model, Eq. 5). These models can be formulated as:

$$\text{ODL model:} \quad Y = \alpha + \beta X + \varepsilon \tag{1}$$

Table 1. Physical properties of orchid seedlings of *Phalaenopsis Sogo F819* at stage 5 (n = 297).

	Area, Y_1 (mm^2)	Height, Y_2 (mm)	Width, Y_3 (mm)	Root length, Y_4 (mm)	Fresh mass, Y_5 (g)	Dry mass, Y_6 (g)
Average	1485	36.0	48.9	28.7	1.8	0.09
SD	819	14.1	19.7	14.5	1.1	0.06
CV	0.55	0.39	0.40	0.51	0.61	0.67

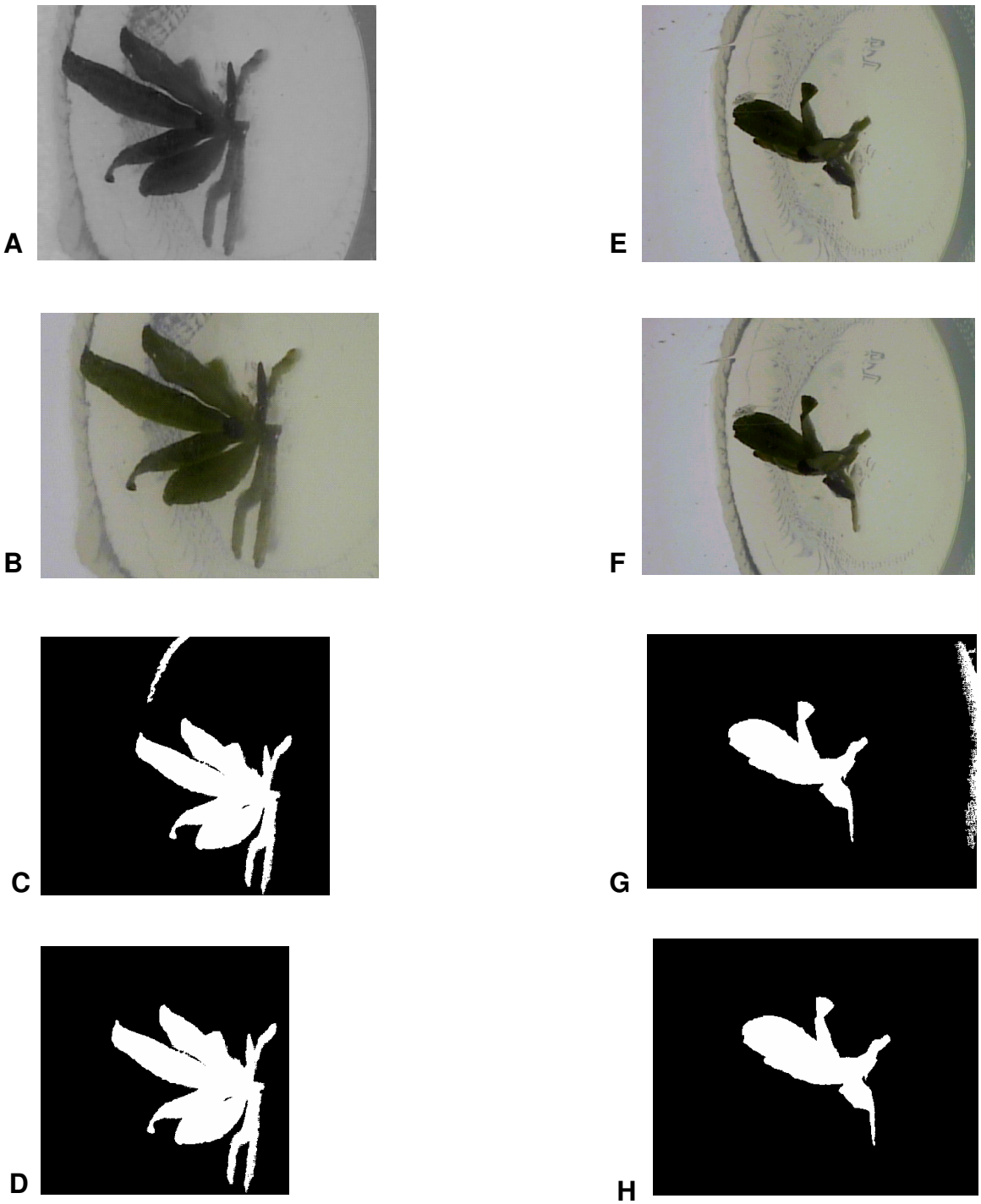

Fig. 2. Orthogonal images of seedling in flask at various processing stages. A, E: color image; B, F: 256 gray-level image; C, G: two-gray-level image; D, H: noise-removed image. A-D: images in long axial direction; E-H: images in short axial direction.

ODP model: $$Y = \alpha + \beta_1 X_1 + \beta_2 X_1^2 + \varepsilon \quad (2)$$

MDL model: $$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \quad (3)$$

EXP model: $$Y = \alpha e^{\beta X} + \varepsilon \quad (4)$$

LN model: $$Y = \alpha + \beta \ln X + \varepsilon \quad (5)$$

where:

X, X_i (i=1..n) = independent variables (i.e., X_1-X_{16} of the image features),

Y = dependent variable (i.e., Y_1-Y_6 of the physical properties),

α, β, β_i (i=1..n) = estimated parameters of model, and

ϵ = error term.

Table 2. Coefficient of linear correlation (*r*) among physical properties (dependent variables) of *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (*n* = 297).

	Area, Y_1	Height, Y_2	Width, Y_3	Root length, Y_4	Fresh mass, Y_5	Dry mass, Y_6
Area, Y_1	1					
Height, Y_2	0.498	1				
Width, Y_3	0.716	0.468	1			
Root length, Y_4	0.646	0.522	0.311	1		
Fresh mass, Y_5	0.959	0.423	0.666	0.647	1	
Dry mass, Y_6	0.952	0.385	0.660	0.622	0.977	1

In terms of linearity of relations between variables, ODL and MDL are linear models while ODP, EXP, and LN are nonlinear models. Nevertheless ODL, ODP, EXP, and LN models are one dimensional models and MDL is a multi-dimensional model. Statistical software, MS-Excel and SAS (V8.02, SAS Institute), were applied in the regression analysis.

Back-propagation neural network (BPN) model

A feed forward back-propagation neural network is one of the supervised neural networks. It consists of an input layer, one or several hidden layers, and an output layer. For a linear application, a hidden layer is not needed. Too many or too few hidden layers or hidden nodes can cause overfitting or underfitting (Næs et al. 1993; Haykin 1994). The BPN is most often applied in the modeling of non-linear and interconnected parameter systems especially in classification and prediction. In general, the classification problem is to separate input vectors into a number of groups that have clear boundaries. On the contrary, the prediction problem deals with giving a continuous value in response to the input vectors. Among these prediction applications, some researchers have focused on agricultural production such as milk yield prediction (Kominakis et al. 2002), output time in wafer fabrication plant forecasting (Chen 2003), and thermal process evaluation parameters for food preservation (Mittal and Zhang 2003).

Basically, the BPN process consists of two passes through the different layers of the network, a forward pass and a backward pass. In the forward pass, an input vector is applied to the sensory nodes of the network, and its effect propagates through the network, layer by layer. Finally a set of outputs is produced as the actual response of the network. During the forward pass, the synaptic weights of the network are all fixed, while during the backward pass, the synaptic weights are all adjusted in accordance with the error correction rule. The synaptic weights are adjusted so as to move the response of the network closer to the desired response. The forward process, backward process, and adjustment of the weights are iterated until the error of the output is satisfactory. Hence the mapping between the input vectors and the output results can be established.

NeuralWorks Professional II/Plus (NeuralWare 1998) software with delta learning rule and sigmoid transfer function was used. The momentum function in NeuralWorks software was also enabled to improve the learning process.

Four BPN models were tested in this research. They are no hidden layer, one hidden layer of two nodes, one hidden layer of four nodes, and two hidden layers of 8-5 nodes. No hidden layer represents a linear model while models with a hidden layer could explore the non-linear relationships between dependent and independent variables. Models with more nodes and more hidden layers could indicate a more complicated relationship between dependent and independent variables. The 16 input nodes are the features obtained from image feature extraction, while one output node was applied to present the physical property. In this investigation, 226 samples were used again in model calibration as training set and the other 71 samples were a testing set.

Prediction error was measured to explore the accuracy of models and was calculated by:

$$\delta = \frac{\sum_{i=1}^{n} \frac{\left|Y_{oi} - Y_{pi}\right|}{Y_{oi}}}{n} \qquad (6)$$

where:

δ = prediction error,
Y_{oi} = observation value,
Y_{pi} = prediction value, and
n = number of observations (n=71).

RESULTS and DISCUSSION

Correlation coefficients among dependent and independent variables

The linear relationships between dependent and independent variables can be evaluated by the coefficient of linear correlation (*r*). Table 2 shows the linear correlation coefficients (*r*) of dependent variables. It indicates that fresh mass to dry mass contained the highest correlation of 0.977, followed by fresh mass to area, and dry mass to area, which were 0.959 and 0.952, respectively. Basically, the plant consisted of water and dry matter when fresh, but when dry it only had dry matter. Samples were chosen at the same culture stage and the same species that maintained similar plant structure resulted in the highest *r* value between fresh mass and dry mass. Because plant growth resulted in an increase in plant mass and area, the *r* values between area and fresh mass or dry mass were also high. Among linear correlation coefficients, root length to width had the lowest correlation, 0.311. Root length is usually affected by plant maturity and environmental conditions, such as the nutrients in the medium, ambient temperature, humidity, etc. Usually farmers let plants grow in height early in the subculture processes, then grow in root at a later stage by adjusting the medium nutrients. The root length is an important criterion for market value. Thus, farmers usually culture plants to get a longer or healthier root in stage 5, where the width of the seedling is not the main concern. This approach may weaken the linearity between root length and plant width.

Table 3. Coefficient of linear correlation (*r*) among image features (independent variables) of *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (*n* = 297).

	X_1*	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_1	1.000	0.798	0.771	0.747	0.861	0.749	0.675	0.683	0.970	0.846	0.792	0.781	0.916	0.794	0.764	0.964
X_2		1.000	0.587	0.784	0.703	0.669	0.661	0.649	0.782	0.909	0.683	0.782	0.741	0.906	0.667	0.778
X_3			1.000	0.794	0.728	0.728	0.670	0.633	0.779	0.722	0.915	0.778	0.654	0.605	0.848	0.776
X_4				1.000	0.687	0.722	0.629	0.667	0.746	0.824	0.779	0.906	0.657	0.757	0.723	0.743
X_5					1.000	0.779	0.751	0.733	0.959	0.812	0.809	0.779	0.925	0.744	0.794	0.965
X_6						1.000	0.583	0.751	0.790	0.917	0.718	0.807	0.700	0.849	0.651	0.788
X_7							1.000	0.789	0.736	0.680	0.912	0.780	0.646	0.588	0.923	0.739
X_8								1.000	0.732	0.767	0.777	0.920	0.658	0.703	0.769	0.733
X_9									1.000	0.861	0.829	0.809	0.954	0.800	0.806	0.999
X_{10}										1.000	0.767	0.870	0.788	0.960	0.721	0.858
X_{11}											1.000	0.852	0.712	0.653	0.969	0.829
X_{12}												1.000	0.720	0.798	0.818	0.808
X_{13}													1.000	0.789	0.736	0.956
X_{14}														1.000	0.644	0.796
X_{15}															1.000	0.808
X_{16}																1.000

* See text for details of variables X_1 to X_{16}.

Table 4. Coefficient of sample determination (R^2) for *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (*n* = 226) for one dimensional linear (ODL) model.

	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6
X_1**	0.822	0.160	0.411	0.423	0.795	0.785
X_2	0.539	0.233	0.278	0.399	0.522	0.521
X_3	0.590	0.263	0565	0.442	0.546	0.532
X_4	0.470	0.252	0.353	0.441	0.442	0.422
X_5	0.837	0.262	0.402	0.439	0.803	0.770
X_6	0.632	0.317	0.482	0.389	0.616	0.579
X_7	0.438	0.387	0.236	0.486	0.407	0.377
X_8	0.550	0.441	0.358	0.522	0.535	0.496
X_9	0.897	0.221	0.440	0.466	0.864	0.843
X_{10}	0.725	0.342	0.474	0.482	0.704	0.679
X_{11}	0.634	0.400	0.474	0.576	0.588	0.560
X_{12}	0.603	0.410	0.416	0.573	0.580	0.543
X_{13}	0.694	0.238	0.418	0.510	0.679	0.676
X_{14}	0.631	0.390	0.321	0.553	0.632	0.586
X_{15}	0.787	0.115	0.273	0.396	0.787	0.770
X_{16}	0.896	0.219	0.437	0.464	0.864	0.841

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass
** See text for the independent variable X_1 - X_{16}.

The *r* value between physical properties suggests the relationship of each set of properties. The sign of *r* indicates the direction of the relationship; positive sign means they have the same trend, i.e., one increases as the other does and vise versa. The magnitude of *r* shows how closely they are correlated. The higher *r* suggests higher correlation between variables. Therefore, the *r* value appears to be useful in analyzing the relationship between variables, and those with higher *r* could be used to predict one another. For example, one could predict fresh mass via a model of dry mass, since fresh mass and dry mass have an *r* of 0.977.

Table 3 shows the *r* values among 16 independent variables. The highest *r* was in variable X_9 (average area of FVI and SVI) and X_{16} (square root of product of FVI and SVI) of 0.999. The higher *r* value suggests that variables between them were highly correlated and sometimes they will cause overfitting in model calibration if they are included in the model simultaneously. The lowest *r* was 0.583 in X_6 (height of SVI) and X_7 (width of SVI). The *r* value for other variables ranged from 0.587 to 0.970. The lower *r* value shows that the variables are not linearly correlated well and offer different information when a calibration model includes them.

An REG procedure supported by SAS calculated the calibration for the MDL model. The procedure is capable of variable selection based on the F-value comparison. The model kept the variables that had high contribution in the calibration. Thus the highly linearly correlated variables will automatically be neglected if one of these variables is already in the model.

Modeling and prediction

Tables 4-6 show the coefficient of sample determination (R^2) between each physical property and the 16 image features in ODL, ODP, and MDL models, respectively. In the ODL model, the highest R^2 occurred with Y_1 (area) with X_9 (average area of FVI and SVI) of 0.897. The R^2 value is an indicator for measuring the degree to which the dependent variable was affected by the independent variable. A higher R^2 means that the independent variable has a strong influence on the changes of the dependent variable. The feature X_9 regularly increased as Y_1 did, thus they obtained a high R^2. Results also show that in area calibration, feature X_9 obtained higher R^2 than feature X_1 (area of FVI) and X_5 (area of SVI) did which had R^2 of 0.822 and 0.877, respectively. This may suggest that the combined feature of FVI and SVI performed better than the single feature of FVI or SVI.

The ODL model is a one dimensional linear model that calibrates dependent variables with only one independent variable. It is a simple form for modeling. Comparing the

Table 5. Coefficient of sample determination (R^2) for *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 226) for one dimensional second order polynomial (ODP) model.

	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6
X_1**	0.830	0.282	0.541	0.435	0.796	0.787
X_2	0.545	0.285	0.340	0.416	0.526	0.523
X_3	0.592	0.289	0.575	0.444	0.551	0.539
X_4	0.473	0.289	0.390	0.444	0.442	0.422
X_5	0.840	0.359	0.518	0.446	0.803	0.770
X_6	0.647	0.378	0.565	0.392	0.618	0.581
X_7	0.463	0.434	0.353	0.486	0.419	0.388
X_8	0.555	0.478	0.460	0.523	0.535	0.496
X_9	0.900	0.349	0.569	0.473	0.864	0.843
X_{10}	0.725	0.412	0.542	0.483	0.707	0.682
X_{11}	0.634	0.423	0.528	0.586	0.592	0.565
X_{12}	0.603	0.440	0.464	0.573	0.592	0.551
X_{13}	0.709	0.328	0.507	0.535	0.683	0.679
X_{14}	0.650	0.461	0.446	0.563	0.633	0.588
X_{15}	0.851	0.200	0.409	0.454	0.832	0.813
X_{16}	0.898	0.344	0.566	0.472	0.864	0.841

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass
** See text for the independent variable X_1 - X_{16}.

Table 6. Coefficient of sample determination (R^2) for *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 226) for multi-dimensional second linear (MDL) model.

	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6
X_6, X_9, X_{15}**	0.902	-	-	-	-	-
X_7, X_8, X_{10}	-	0.517	-	-	-	-
X_3, X_6, X_{14}	-	-	0.602	-	-	-
X_3, X_{11}, X_{12}, X_{14}	-	-	-	0.626	-	-
X_6, X_9, X_{13}, X_{14}	-	-	-	-	0.878	-
X_9, X_{14}	-	-	-	-	-	0.854

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass
** See text for the independent variable X_1 - X_{16}.

Table 7. The highest coefficient of sample determination (R^2) of Y_1 (area) for various models for *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 226).

	ODL	ODP	MDL	EXP	LN
Y_1 to X_9	0.897	0.900	N/A	0.707	0.743
Y_1 to X_6, X_9, X_{15}	N/A	N/A	0.902	N/A	N/A

ODL: One dimensional linear model
ODP: one dimensional second order polynomial model
MDL: Multi-dimensional linear model
EXP: Exponential model
LN: Natural log model

dependent variables, Y_1 obtained the highest R^2 followed by Y_5 (fresh mass) and Y_6 (dry mass). The independent variable contribution to area calibration models was X_9, the same as for Y_5 and Y_6. This situation may cause a high linear correlation between Y_1, Y_5, and Y_6 that had an r of over 0.95. The highest R^2 for dependent Y_2 (height) was calibrated by X_8 (circumference of SVI). The height of seedling will increase as the plant grows and also will lengthen its leaf, resulted in increasing of X_8. In width (Y_3) the highest R^2 was obtained with X_3 (width of FVI) which suggested that the front view of the image performed better in width measurement than the side view image.

Table 5 shows the R^2 value for the dependent and independent variables in the ODP model. Instead of a linear relation, a nonlinear second order polynomial was applied and its results showed that area (Y_1) outperformed the highest R^2 of 0.900 followed by fresh mass (Y_5) and dry mass (Y_6) that had R^2 of 0.864 and 0.843. In comparison with the ODL and ODP models, the ODP performed better in Y_1, Y_2, Y_3, and Y_4 in terms of R^2, but they only had the same R^2 for Y_5 and Y_6. In Table 5, the R^2 of dependent variables to independent variables X_9 and X_{16} were very close. For instance, R^2 was 0.90 for Y_1 to X_9 and 0.898 for Y_1 to X_{16}. This might result from the highly linear correlation between independent variable X_9 and X_{16} that had an r of 0.999.

The R^2 for MDL models when the independent variables were added to calibrate a multidimensional linear model is shown in Table 6. Compared to ODL and ODP models, the R^2 was generally improved in the MDL model by an average of 4.6 and 3.0%, respectively. The independent variables were selected by SAS REG procedure which tested every independent variable in the model calibration and kept the variables that made a significant contribution. Among MDL models, the highest R^2 occurred on Y_1 to X_6, X_9, and X_{15}. The highly correlated independent variables X_9 and X_{16} were not both included in the model. The r value among X_6, X_9, and X_{15} ranged from 0.651 to 0.806 (shown in Table 2).

Two other one-dimensional models named EXP and LN were also tested and the highest R^2 values were listed and compared in Table 7. The R^2 for Y_1 were only 0.707 and 0.743 for EXP and LN model and were about 21 and 14% less compared to ODL and ODP models.

Among five regression models, the MDL model had the highest R^2 value for the six dependent variables. Each equation suggested by the MDL calibration process was as follows:

$$Y_1 = 1.45X_6 + 32.73X_9 - 3.37X_{15} - 0.78 \quad (7)$$

$$Y_2 = 13.15X_7 + 4.18X_8 + 9.34X_{10} + 9.83 \quad (8)$$

$$Y_3 = 28.72X_3 + 13.14X_6 - 0.22X_{14} + 0.47 \quad (9)$$

$$Y_4 = -4.27X_3 + 24.3X_{11} + 7.8X_{12} + 2.02X_{14} - 2.87 \quad (10)$$

$$Y_5 = 0.145X_6 + 3.65X_9 - 0.04X_{13} + 0.133X_{14} - 0.186 \quad (11)$$

$$Y_6 = 0.187X_9 + 0.009X_{14} - 0.01 \quad (12)$$

Table 8. Prediction error of six physical properties by MDL (multi-dimensional linear) model (n = 71).

	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6
X_6, X_9, X_{15}**	0.257	-	-	-	-	-
X_7, X_8, X_{10}	-	0.323	-	-	-	-
X_3, X_6, X_{14}	-	-	0.159	-	-	-
X_3, X_{11}, X_{12}, X_{14}	-	-	-	0.533	-	-
X_6, X_9, X_{13}, X_{14}	-	-	-	-	0.542	-
X_9, X_{14}	-	-	-	-	-	0.592

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass
** See text for the independent variable X_1 - X_{16}.

the 226 training samples. The weights of the nodes were adjusted during the training epoch to meet the relationships between dependent and independent variables. Among these models, the two hidden layers with eight nodes in the first layer and five nodes in the second layer showed the lowest prediction error, followed by one hidden layer of two nodes. The linear network had a relatively high error. This suggested that relationships among physical properties and the 16 image features could be correlated with the nonlinear network to improve prediction accuracy.

Applying this two hidden layers BPN model to predict the six physical properties we obtained the

Table 9. Prediction error of various neural network models on seedling area (Y_1) for *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 71).

	No hidden layer	One hidden layer of 2 nodes	One hidden layer of 4 nodes	Two hidden layers of 8-5 nodes
Prediction error	0.192	0.120	0.124	0.118

Table 10. Prediction error of neural network model with two hidden layers of 8-5 nodes for the six physical properties of *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 71).

Physical properties	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6
Prediction error	0.118	0.265	0.186	0.481	0.520	0.581

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass

Table 11. Comparison of prediction error between neural network model of two hidden layers of 8-5 nodes (K_1) and MDL (multi-dimensional linear) model (K_2) for the six physical properties of *Phalaenopsis Sogo Vivien F819* seedlings at stage 5 (n = 71).

Physical properties	Y_1*	Y_2	Y_3	Y_4	Y_5	Y_6	Average
Difference (K_1 - K_2)	-0.139	-0.058	0.027	-0.052	-0.22	-0.011	-0.043
Improved ratio $\frac{K_1 - K_2}{K_2}$	54.1%	18.0%	17.0%	9.7%	4.1%	1.9%	10.7%

* Y_1: area Y_2: height Y_3: width
Y_4: root length Y_5: fresh mass Y_6: dry mass

The prediction errors for Eqs. 7 to 12 are shown in Table 8; they indicate that Y_3 had the lowest prediction error of 0.159 followed by Y_1 of 0.257 and the overall prediction error averaged 0.401. This result suggested the highest R^2 did not necessarily reach lowest prediction error because some models may be overcalibrated. For instance, the R^2 for Y_1 was 0.902 and for Y_3 was 0.602. The reasons for these miss-predictions may be the variance of plants, the image features, and the models of prediction.

BPN modeling and prediction

The results of prediction error on 71 samples in the Y_1 variable for various BNP models are shown in Table 9. Each model tabulated in Table 9 had been trained for 200,000 epoch with

results shown in Table 10. These show that the prediction error ranged form 0.118 to 0.581. The lowest prediction error was on Y_1 (area) and the highest prediction error was on Y_6 (dry mass). Comparing the prediction error between the two hidden layers BPN model and MDL model, the difference and improved rate of the BPN model are presented in Table 11. The BPN model generally performed better than the MDL. The average improvement rate was 10.7% and the area contained the highest improved rate of 54.1%. The fact that the BPN model calibrated the relationship by every independent variable and adjusted their influences by changing their synaptic weights may contribute to this improvement.

The prediction error in the two layers BPN model and MDL model for Y_4 (root length), Y_5 (fresh mass), and Y_6 (dry mass) was relatively higher than other dependent variables. This might result from variance of the seedlings or the calibration model was an overfitting model. Further study to improve the prediction accuracy and model numbers seedlings in one flask is suggested to extend its application.

CONCLUSIONS

This study applied a back lighting machine vision system for measurement of flask seedlings of orchid Phalaenopsis. The measurement models calibrating six physical properties with 16 image features were explored by linear and nonlinear regression and back propagation neural network (BPN). Based on the results, we conclude:

1. Among five regression models, the multi-dimensional linear model (MDL) performed best in the calibration process in terms of sample coefficient of determination (R^2) that had R^2 of 0.902 for plant area (Y_1). Applying MDL to predict 71 samples, the average prediction error was 0.401 and the lowest prediction error occurred with plant width (Y_3) of 0.159.
2. The prediction error for the two hidden layers of 8-5 nodes BPN model averaged was 0.359 on six properties and the model performed about 10% better than the MDL model. The lowest prediction error occurred for the plant area (Y_1) of 0.118 and it indicated prediction accuracy of about 88.2%. Using weight adjustment to relate physical property and image features may contribute to this improvement of prediction accuracy. Nevertheless, the lowest accuracy was on dry mass (Y_6). The variance of dry mass and the features selected in the model may diminish its prediction accuracy. The findings of this study might help farmers in evaluating their products and ensure the quality of flask seedlings in stage 5.
3. Application of other image features to improve the prediction accuracy and study the properties prediction for the format of many seedlings in one flask is recommended.

ACKNOWLEDGEMENT

The authors thank SOGO Orchids Nursery Company and Professor C.H. Hsieh for their full support on material supply and technical suggestions. The authors also thank Dr. Y.R. Chen, Director of ISL, USDA, for his constructive help in Neural Network analysis.

REFERENCES

ASAE. 2001. ASAE S487 - Moisture content measurement for tobacco. In *ASAE Standards 2001*, 599. St. Josephs, MI: ASAE.

Britt, J. 2000. The status of the commercial production of potted orchid around the world. *ASHS-2000 Symposium: Potted Orchid Production in the New Millennium*. http://primera.tamu.edu/orchids/britt.htm (2004/12/17)

Chen, C.C. 2002. Introduction on orchid Industry. http://bse.nchu.edu.tw/services.htm (in Chinese) (2004/12/17)

Chen, C.C. and R.S. Lin. 1999. Nondestructive dry-matter estimation of phalenopsis leaves. *Journal of Chinese Meteorology* 6(2): 87-96. (in Chinese)

Chen, I. 2004. Management and utilization of intelligent right of orchid *Phalaenopsis.* Monthly Report of Taiwan Floriculture 202: 46-48. (in Chinese)

Chen T. 2003. A fuzzy back propagation network for output time prediction in a wafer fab. *Applied Soft Computing* 2/3F: 211-222.

Chen X, S.M. Welch, Z.N. Zhang and D. Armbrust. 2001. Measurement of change in soybean plant cross-sectional area under wind conditions via image processing. *Transactions of the ASAE* 44(6): 1923-1929.

Chen Y.R., K. Chao and M.S. Kim. 2002. Machine vision technology for agricultural applications. *Computers and Electronics in Agriculture* 36: 173-191.

Griesbach, R.J. 2000. Potted phalaenopsis orchid production: History, present status, and challenges for the future. *ASHS-2000 Symposium: Potted Orchid Production in the New Millennium*. http://primera.tamu.edu/orchids/griesbach.htm (2004/12/17)

Huang, K.J. and T.C. Lin. 2000. Estimating the geometric characteristics of Phalaenopsis orchid during big plant stage with machine vision. *Journal of Agricultural Machinery* 9(2):13-26. (in Chinese)

Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation,* 138-235. New York, NY: Macmillan College Publishing.

Hsieh, C.L., S.F. Jeng and T.T. Lin. 1997. Application of image texture analysis and neural network on the growth stage recognition for head cabbage seedlings. *Journal of Agricultural Machinery* 6(2): 1-13. (in Chinese)

IMAQ. 1999. IMAQ *Vision for G Reference Manual*, June 1999 edition. Austin, TX: National Instruments.

Jeng, S.F. and T.T. Lin. 1997. Growth measurement and modeling of cabbage seedlings (1)--Implementation of the automatic measurement system. *Journal of Chinese Agricultural Machinery* 6(4): 69-82. (in Chinese)

Kominakis A.P., Z. Abz, I. Maltaris and E. Rogdakis. 2002. A preliminary study of the application of artificial neural networks to prediction of milk yield in dairy sheep. *Computers and Electronics in Agriculture* 35: 35-48.

Laws, N. 2004. Orchid commerce around the world. *FloraCulture International.* http://www.floraculturein tl.com/archive/display.asp?ArticleID=293 (2004/12/17)

Lin, P.P. and V.N. Ruzhitsky. 1996. Machine vision techniques for measuring the canopy of tomato seedling. *Journal of Agricultural Engineering Research* 65: 85-95.

Lin, T.T., W.C. Liao and C.F. Chien. 2001. 3D graphical modeling of vegetable seedlings based on a stereo machine vision system. ASAE Paper No. 013137. St. Joseph MI: ASAE.

Mittal, G.S. and J. Zhang. 2002. Prediction of food thermal process evaluation parameters using neural networks. *International Journal of Food Microbiology* 79: 153-159.

NeuralWare. 1998. NeuralWorks Professional II/Plus. Carnegie, PA: NeuralWare, Inc.

Næs, T., K. Kvaal, T. Isaksson and C. Miller. 1993. Artificial neural networks in multivariate calibration. *Journal of Near Infrared Spectroscopy* 1:1-11.

Perry, L. 2003. Orchids for your valentine. *The Green Mountain Gardener*. http://pss.uvm.edu/ppp/articles/orchids/html (2004/12/17)

SAS. 1999. SAS system for windows, release 8.0. Cary, NC: SAS Institute Inc.

Thomas V., P. Treitz, D. Jelinski, J. Miller, P. Lafleur and J.H. McCaughey. 2002. Image classification of a northern peatland complex using spectral and plant community data. *Remote Sensing of Environment* 84: 83-99.

Visen, N.S., J. Paliwal, D.S. Jayas and N.D.G. White. 2004. Image analysis of bulk grain samples using neural networks. *Canadian Biosystems Engineering* 46:7.11-7.15.

Yang, C.C., S.O. Prasher and J.A. Landry. 2002. Weed recognition in corn field using back-propgation neural network models. *Canadian Biosystems Engineering* 44: 7.15-7.22

LIST of SYMBOLS

EXP	exponential model
FVI	Front View Image
MDL	multi-dimensional linear
NN	neural network
ODL	one dimensional linear
ODP	one dimensional second order polynomial
SVI	Side View Image
Y_1	area (mm^2)
Y_2	height (mm)
Y_3	width (mm)
Y_4	root length (mm)
Y_5	fresh mass (g)
Y_6	dry mass (g)
X_1	area of FVI (Front View Image)
X_2	height of FVI
X_3	width of FVI
X_4	circumference of FVI
X_5	area of SVI (Side View Image)
X_6	height of SVI
X_7	width of SVI
X_8	circumference of SVI
X_9	average area of FVI and SVI
X_{10}	average height of FVI and SVI
X_{11}	average width of FVI and SVI
X_{12}	average circumference of FVI and SVI
X_{13}	number of pixels product of FVI and SVI
X_{14}	width product of FVI and SVI
X_{15}	height product of FVI and SVI
X_{16}	square root of product of FVI and SVI
δ	prediction error