

# Decision Support System Using Artificial Neural Network to Predict Rice Production in Phimai District, Thailand

Saisunee Jabjone and Sura Wannasang

**Abstract**—In Thailand, nowadays, the planted area, climate and rainfall are changed rapidly effect to unstable of Thai rice production. The decision-making processes often require reliable rice response models. Local governor and farmer need simple and accurate estimation techniques to predict rice yield in the planning process. This study aims to develop the decision support system using Artificial Neural Networks (ANN) by adjust the value of parameters and study about 9 Algorithms training. In predicting rice productions which its study found that each values that was adjusted making high predicting like appropriate number of hidden nodes to model equals to 9, learning rate effects the speed of appropriate learning to the model equals to 0.5, and appropriate momentum to model was 0.5. CGB Algorithm has coefficient decision higher than using regression variable technique by Stepwise multiple method ( $R^2=0.9884$  and  $R^2=0.734$ ). In addition, the area under the ROC curve of ANN and stepwise multiple regression method was 4,293.70 and 40,160.00, respectively.

**Index Terms**—Prediction, decision support system, artificial neural networks, stepwise multiple regression method.

## I. INTRODUCTION

Rice production has long played a vital role in Thailand's socio-economic development, making the country the world's largest rice exporter in the last 3 decades. Although Thailand has had the lowest yield among the world's top 10 rice-producing countries, even lower than the world average, Thailand has been the largest rice-exporting country for almost 30 years now [1]. This is because of the different varieties of rice produced among the top 10 rice-producing countries, with some countries utilizing more new high-yielding varieties while the others grow low-yielding local varieties. Although the Thai national average yield is low, the total production is much higher than the domestic utilization, creating at least 6–9 million tons of rice surpluses for export each year. Rice production area in Thailand has a total of 9.9 million ha (6.2 rai = 1 ha) in 2011 with total rice yield production of 34.5 million ton [2]. Even irrigation can support only approximately 25% of the total rice area, together with appropriate varieties can produce higher average rice yield compared with those of rainfall area [3].

The ability to predict the future enables the farmer to take the most appropriate decision in anticipation of that future. An

artificial neural network (ANN) offers exciting possibilities to perform machine learning and prediction, and abundantly utilized in performing agriculture prediction task [4], [5]. An artificial neural network is a computational mechanism that is able to acquire, represent, and compute a weighting or mapping from one multivariate space of information to another, given a set of data represent on that mapping [6]. It can identify subtle patterns in input training data which may be missed by conventional statistical analysis. In contrast to regression models, neural networks do not require knowledge of the functional relationships between the input and the output variables [7]. Researchers from many scientific disciplines are designing artificial neural networks to solve a variety of problems in pattern recognition, prediction, optimization, associative memory, and control. Many researchers have developed several predicting crop yield models in relation to different parameters as influencing factors by applications of artificial neural networks and by combining ANN and statistical techniques such as linear regression technique. Puteh Saad *et al.*, used Backpropagation Network to predict rice yield using climatic observation data and predicted with a maximum of 45-60kg/ha [8]. B. Ji *et al.*, used neural network to predict rice yield based on soil parameters and achieve a testing error of 17.3% [9]. Liu *et al.*, used ANN to predict maize yield based on rainfall, soil and other parameters and obtained a testing error of 14.8%, whereas O'Neal *et al.*, 2002 used Backpropagation Network to predict rice yield based on weather data [10], [11]. Jin *et al.* used a neural network combined with the mean generating function (MGF) to build a hybrid prediction model and conducted a predict experiment on the precipitation in the northern, central, and southern regions of the Guangxi province [12]. Huang *et al.* conducted research on a principal-component-based neural network model and applied it to a water level prediction [13].

This study therefore aims to develop the decision support systems for rice production using artificial neural networks to create a prediction model. The study area is in Phimai district. It locates in Nakhon Ratchasima province, northeastern of Thailand. It is about 300 kilometers from Bangkok. This district is subdivided into 12 subdistricts (tambon), which are further subdivided into 208 villages (mooban). The number of population is 129,849 which majority of population are farmers. The paddy rice has about 3.8 million rais. The rice yield for major rice cultivation and second rice cultivation were 285kilograms/rais and 640kilogram/rais, respectively. In Phimai, therefore, second rice cultivation has more paddy yield than major rice cultivation because of the irrigation system.

This study utilized the data published by the Ministry of

Manuscript received October 5, 2013; revised December 8, 2013. This work was supported in part by Nakhon Ratchasima Rajabhat University, Thailand.

Saisunee Jabjone and Sura Wannasang are with the Nakhon Ratchasima Rajabhat University, Thailand, 30000. (e-mail: a1102923@hotmail.com, sura13@hotmail.com).

Agriculture and Cooperatives of Thailand in 2010 to assess rice production in Phimai district. It was to develop a simple technique for farmers and local governor for predicting rice production in different subdistrict of Phimai. In addition, this work will partly assist local governor in analyzing and assessing data needs for various purposes.

## II. MATERIALS AND METHODS

The advances in computing and information storage have provided the extract knowledge from the raw data. It has lead to new methods and techniques such as data mining that can bridge the knowledge of the data to the rice yield prediction. A neural network is a powerful data modeling tool that is able to capture and represent complex input/output relationships. The motivation for the development of neural network

technology stemmed from the desire to develop an artificial system that could perform "intelligent" tasks similar to those performed by the human brain [14]. An Artificial Neural Network (ANN) is one of mathematical method for modeling of different phenomenon [15]. The ANN can be applied to solve problems involving complex relationships between variables [16]. It can learn new patterns that are not previously available in the trained data sets, and since it can update knowledge over time as long as more training data sets are provided [17], therefore, it is expected that ANN could simulate the rice production using the effective parameters. In this study, we used ANN to build the prediction model. The performance of ANN model was compared with Statistical Method (Stepwise multiple regression method). The process of build of the prediction model for rice production is divided into four steps as show in Fig. 1.

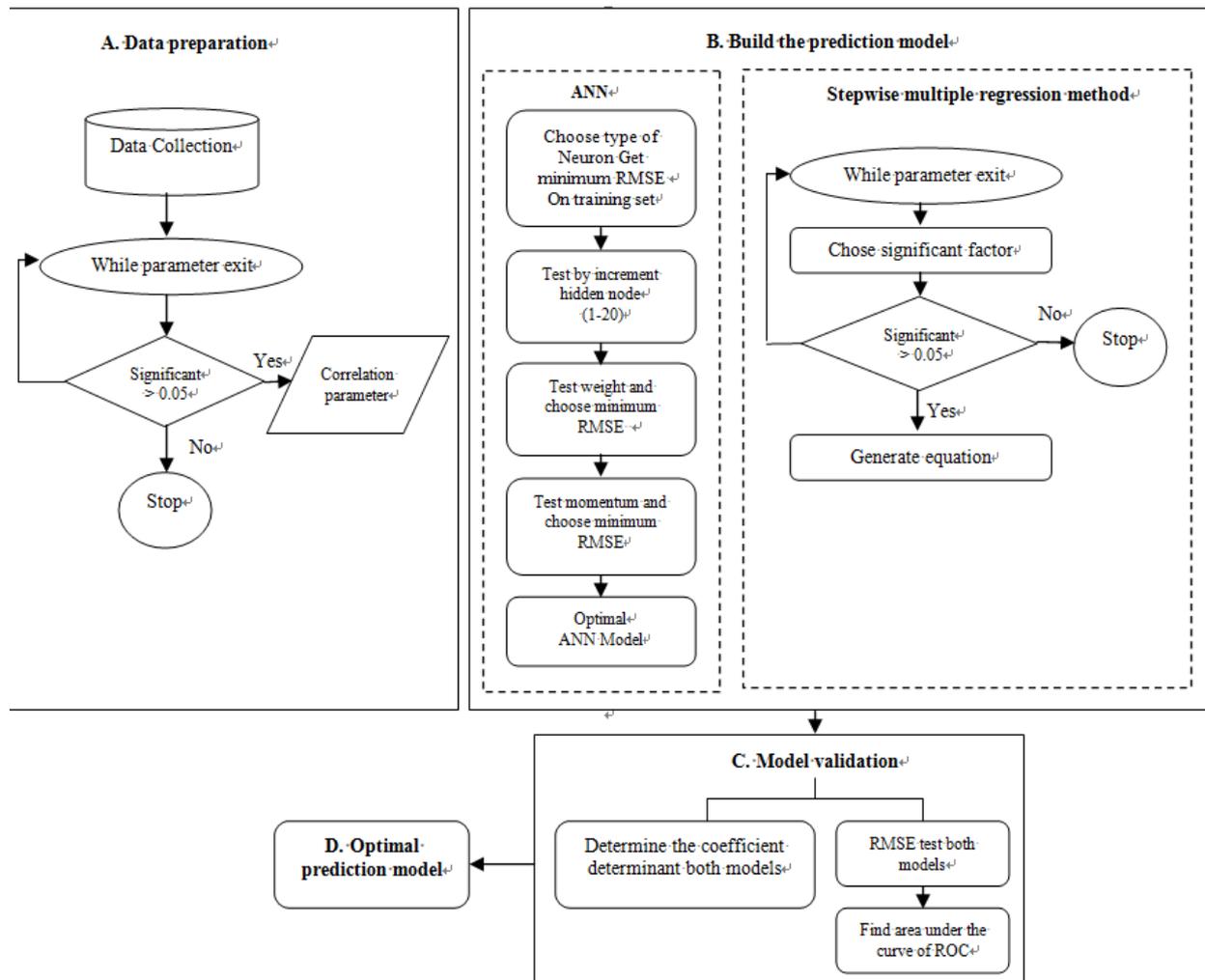


Fig. 1. Process of developing the prediction model

### A. Data Preparation

The rice cultivation data collected through several organizations in Thailand such as Ministry of Agriculture and Cooperatives, Thai Meteorological Department, Land Development Department and Royal Irrigation Department. In addition, the volunteers from agriculture of Phimai district interviewed the local farmer during their field visits about rice yield, type of rice cultivation, effect of disaster from flood and rainless. The data were collected for 5 year from 2008 to 2012.

The data came from 12 subdistricts (208 villages) with two planting season for each year, in season and dry season. These data warehouse were implemented on SQL server 2008. Each of these data marts had information related to statistics of rice yield, rice breed, irrigated area, cultivated area, planted damage area, rice field characteristic (low land, high land), and season (in season, dry season). The validation checks were implemented wherever possible to provide the quality data for the end users. These validation checks were of different forms such as detections of outliers, consistency

check of the data flow and its aggregations from lower level to upper level in a hierarchy, aggregation check within the same level of hierarchy, etc. These collected data then were separated into training and test sets. 60% of the data were utilized for training the predictive model and 40% of the data were reserved for testing. Out of 1,648 total set data, 1,401 sets were chosen for training and 247 set were used for prediction. The Pearson Product-Moment Correlation coefficient was used to find the correlation of parameters for using in predictive model. The equation is shown in (1).

$$r_{xy} = \frac{\sum (X_i Y_i) - (\sum X_i)(\sum Y_i) / n}{\sqrt{\left[ \sum X_i^2 - \left( \frac{\sum X_i}{n} \right)^2 \right] \left[ \sum Y_i^2 - \left( \frac{\sum Y_i}{n} \right)^2 \right]}} \quad (1)$$

$r$  is the value to measure linear of variable between  $X$  and  $Y$ . If the least significant is more than 0.05, it means the variables  $X$  and  $Y$  are correlated relationship.

### B. Build the Prediction Model

In this study, we compared two techniques, Artificial Neural Network and Stepwise multiple regression method, for finding the optimal prediction model.

- Using Artificial Neural Network (ANN) to build the prediction model, it separated into 3 steps as show in Fig. 1. *Step 1*, In order to find out an optimal configuration of neural network model, it was necessary to combine many different ANN prototypes. Consequently, nine different training algorithms were used for network training including Levenberg-Marquardt: LM, BFGS Quasi-Newton: BFG, Resilient Backpropagation: RP, Scaled Conjugate Gradient: SCG, Conjugate Gradient with Powell/Beale Restart: CGB, Fletcher-Powell Conjugate Gradient: CGF, Polak-Ribière Conjugate Gradient: CGP, One Step Secant : OSS and Variable Learning Rate Backpropagation: GDX. *Step 2*, the number of hidden layer(s) and the value of the training parameters for every training algorithm were obtained by trial and error method with considering of ANN performance. For selecting the number of hidden layers along with the right number of neurons in the middle layers, comparison of networks which had different number of neurons and also different number of hidden layers were carried out. Comparison of the performances of the developed ANN models was conducted based on the scale of Root Mean Square Error (RMSE). To measure the impact of the number of nodes in the head, this study changed the number of nodes from 1 to 20 nodes. The optimal model was determined by minimum value of RMSE. Finally, *Step 3*, It was the initialization of the network weights and parameters by adjust the momentum.
- The statistical method (Stepwise multiple regression method: SMR) was used to compare with the ANN prediction model. Stepwise multiple is the method to determine parametric equation by linear algebra using invert matrix (2).

$$\begin{pmatrix} a_{11} & \mathbf{K} & a_{1n} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ a_{m1} & \mathbf{L} & a_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_1 \\ b_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_1 \\ r_n \end{pmatrix} \quad (2)$$

$$\begin{pmatrix} b_1 \\ b_1 \\ b_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_1 \\ r_n \end{pmatrix} \begin{pmatrix} a_{11} & \mathbf{K} & a_{1n} \\ \mathbf{M} & \mathbf{O} & \mathbf{M} \\ a_{m1} & \mathbf{L} & a_{mn} \end{pmatrix}^{-1}$$

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_n X_n$$

### C. Model Validation

To compare the performance of prediction models between ANN and SMR, they were conducted based on the scale of Root Mean Square Error (RMSE) and coefficient of determination ( $R^2$ ). The equation is shown in (3) and (4).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - F_i)^2}{n}} \quad (3)$$

$$R^2 = \frac{(\sum XY - N\bar{X}\bar{Y})^2}{(\sum X^2 - N\bar{X}^2)(\sum Y^2 - N\bar{Y}^2)} \quad (4)$$

### D. Choose the Optimal Prediction Model

The area under the curve of ROC (5) is used to compare the performance between prediction model of ANN and SMR.

$$Area = \int_0^n F_{rmse}(x) dx = \lim_{\Delta x \rightarrow 0} \sum_{i=0}^{n-1} (F_{rmse}(x_i) + F_{rmse}(x_{i+1})) \frac{\Delta x}{2} \quad (5)$$

## III. RESULTS AND DISCUSSION

In this study, MATLAB was used to build the prediction model. The result divided into the following topics;

### A. Data Correlation

To find the correlation of parameters for using in predictive model, they were 7 factors that effected to the rice production. It consisted of irrigation zone, rice field characteristic, season, rice breed, cultivation methods, cultivated area, and damage area.

### B. Prediction Model

- This study, we compared the prediction model of 9 learning algorithms. The Conjugate Gradient with Powell/Beale Restarts: CGB provided the best RMSE value (0.0061) as show in Fig. 2.

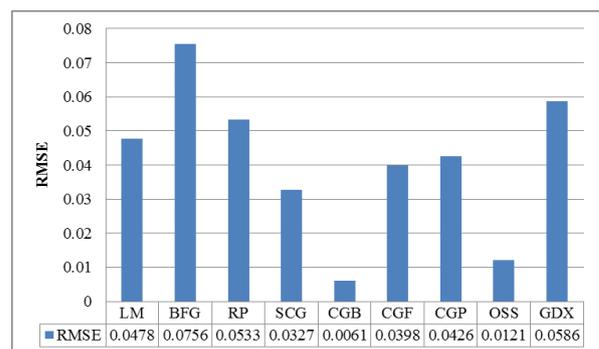


Fig. 2. RMSE of nine learning algorithm

The optimal hidden node of CGB algorithms was node 9

which gave the best value of RMSE (0.0838). The training model was adjusted by changing the momentum rate from 0 to 1, increment by 0.1. The minimum RMSE was 0.869 at momentum 0.5 (see in Fig. 3). The RMSE rate was represented by quadratic form show in (6).

$$y = 0.27x^2 - 0.27x + 0.94 \quad (6)$$

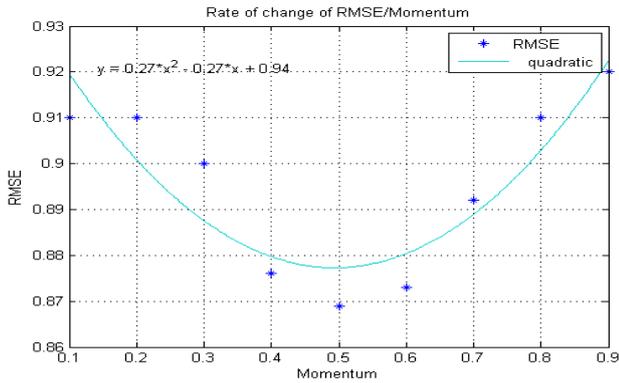


Fig. 3. RMSE by changing the momentum rate

In addition, the coefficient determinant ( $R^2$ ) between the estimate and actual data was 0.988. In Fig. 4, the graph showed the regression line using CGB learning algorithms gave the estimate value closed to the actual data.

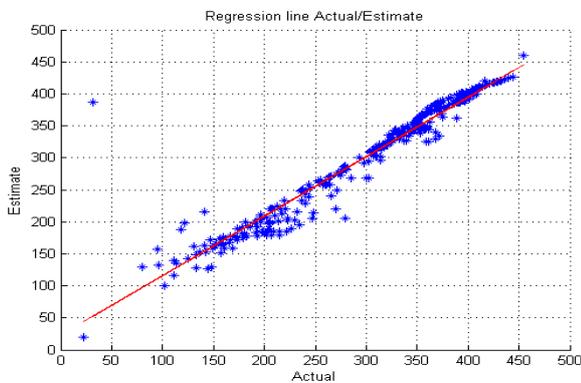


Fig. 4. Regression line between actual data and estimate value using CGB learning algorithms

2) The statistic method (Stepwise multiple regression method: SMR) was used to compare the output with the ANN model. In the SMR model, there were four factors and four equations that effected to the rice production. The coefficient determinant ( $R^2$ ) and SMR equations were shown in Table I.

TABLE I: SMR EQUATION AND  $R^2$

Factor	Equation	$R^2$
A	$Y = 0.48 + 0.31A$	0.836
A+D	$Y = 0.525 + 0.325A - 0.212D$	0.853
A+D+FC	$Y = 1.503 + 0.325A - 0.212D - 0.843FC$	0.854
A+D+FC+I	$Y = 2.828 + 0.325A - 0.212D - 1.150FC - 0.635I$	0.875

A = cultivated area

D = planted damage area

FC = rice field characteristic; 0: lowland and 1: highland

I = irrigated area; 0: none and 1: irrigation

The last equation gave the best value of  $R^2$  (0.875). To test the performance of this equation, the coefficient determinant

( $R^2$ ) between the estimate and actual data was tested. The value of  $R^2$  was 0.734. The regression line between actual and estimate data is shown in Fig. 5.

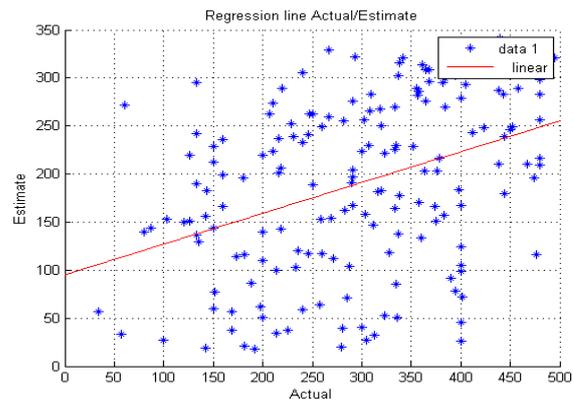


Fig. 5. Regression line between actual data and estimate value using stepwise multiple regression method

### C. Compare the Prediction Model

The ROC curve of RMSE was used to compare the error of both prediction models. These values of ANN and SMR were 4,293.70 and 40,160.00, respectively (see in Fig. 6).

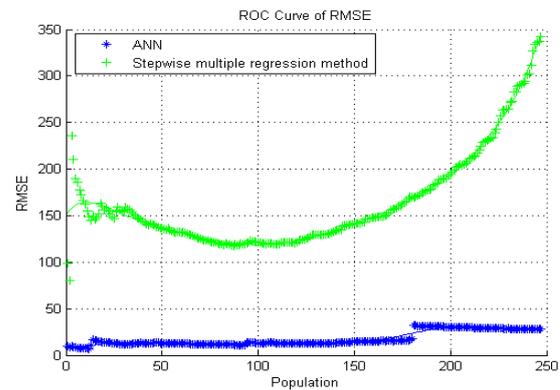


Fig. 6. ROC curve of RMSE between ANN and SMR

### D. Optimal Prediction Model

The area under the curve of ROC in ANN model was less than SMR model. It means that the prediction model using Artificial Neural Network is optimal to predict the rice production.

## IV. CONCLUSION

Rice is the world's most important food crop and a primary source of food for more than half of the world's population. Almost 90% of rice is produced and consumed in Asia, and 96% in developing countries. Raising level of national rice self-sufficiency has become a strategic issue in the agricultural ministry of Thailand. The numerous problem associated with rice farming include monitoring the maintaining irrigation infrastructures, obtaining quality seedlings, and many other problems that need to be addressed in order to increase productivity. These problems can be overcome with a good prediction model that can predict rice yield given the seven parameters. In this study, the optimal prediction model was conjugate Gradient with Powell/Beale Restarts: CGB which the number of hidden nodes that fit to

the model is 9. Learning rate affects the speed of learning appropriate to the model was 0.5 and the momentum that appropriate to the model was 0.5. This artificial neural networks model was compared to multi variable regression analysis by using Stepwise multiple. ANN had the area under the curve of RMSE less than prediction output of Stepwise Multiple Regression Method. We can describe the decision making coefficient value of ANN better than multi variable regression analysis by using Stepwise multiple regression method. It should be possible because variable is used both quantitative and qualitative which artificial neural network will be better in explanation.

#### ACKNOWLEDGMENT

This research was made possible by the funding of Nakhon Ratchasima Rajabhat University and Research and Development Institute. I would like to thank Mrs. Parichart who provided raw data and contacted to district governor. Thank you to officer of Ministry of Agriculture and Cooperatives, Thai Meteorological Department, Land Development Department and Royal Irrigation Department, who supported the data, gave the advice and tested the system.

#### REFERENCES

[1] S. Supunnabul, "Rice production and marketing," (in Thai). *Office of Agricultural Economics Journal*, vol. 52, pp. 53-60, Dec 2012.

[2] J. Apichat, "Crop cutting yield per rai Second rice (Dry season), year 2012, (officers manual)," *Office of Agricultural Manual Report*, Bangkok, Mar 2013.

[3] A. Sayamwara and W. Na Ranong, *Compilation of Rice Knowledge*, Bangkok: Thailand Development Research Institute, 2010.

[4] G. Zhange, "Predicting with artificial neural network," *International Journal of Predicting*, vol. 14, pp.35-62, Jan 1998.

[5] E. Patuwo, M. Y. Hu, and M. S. Hung, "Two-group classification using neural networks," *Decision Science*, vol. 24, no. 4, pp.825-845, July 1993.

[6] S. Haykin, *Neural Networks A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[7] E. Turban, *Decision Support System and Expert Systems*, New Jersey, USA: Prentice Hall, 1998, pp. 694,

[8] S. Puteh, M. Rizon, M. Juhari, J.Nor Khairah, S. Siti Kamarudin, B. Aryati, and R. Nursalasawati, "Backpropagation algorithm for rice

yield prediction," in *Proc. 9th of the Ninth Int. Symp. on Artificial Life and Robotics*, Beppu, Japan, Oita, 2004, pp. 586-589.

[9] B. Ji, Y. Sun, S. Yang, and J. Wan, "Artificial neural networks for rice yield prediction in mountainous regions," *Journal of Agricultural Science*, vol. 145, pp. 249-261, May 2007.

[10] J. Liu, C. E. Goering, and L. Tian, "A neural network for setting target corn yields," *Transaction of the ASAE*, vol. 44, no. 3, pp. 705-713, July 2001.

[11] M. R. O'Neal, B. A. Engel, D. R. Ess, and J. R. Frankenberger, "Neural network prediction of maize yield using alternative data coding algorithms," *Biosystems Engineering*, vol. 83, no.1, pp. 31-45, May 2002.

[12] L. Jim, G. Randolph, and M. Mary, "On the construction and training of reformulated radial basis function neural networks," *IEEE Trans. on Neural Networks*, vol. 14, July 2003.

[13] K. Nakornphanom, C. Lursinsap, and A. Rugchatjaroen, "Fault immunization model for elliptic radial basis function neuron," in *Proc. 9th International Conference on neural Information Processing 2002*, pp.1027-1031.

[14] S. T. Drummond and A. Joshi, "Predictive ability of neural networks for site-specific yield estimation," in *Proc. Second International Geospatial Information in Agriculture and Forestry Conference*, Lake Buena Visat, Florida, Jan 2000, pp. 125-129.

[15] S. Puteh, "Trademark image classification approaches using neural network and rough set theory," Ph.D. dissertation, University Technology Malaysia, 2003.

[16] L. H. Tsoukalas and R. E. Uhrig, *Fuzzy and Neural Approaches in Engineering*, New York: John Wiley & Sons, 1997.

[17] J. P. Bigus, *Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support*, New York: McGraw-Hill, 1996.



**Saisunee Jabjonee** was born in March, 1972. She graduated Doctoral degree from the University of Technology, Sydney, Australia, major in Computing. She is an Assistant Professor at the department of Informatics, Nakhon Ratchasima Rajabhat University, Thailand. Her research interest is in the area of data mining, ontology, data management, database, and education technology.



**Sura Wannasang** was born in May, 1974. He graduated Master of Information Technology, major in software engineering, from KhonKaen. He works as a lecturer at the department of Informatics, Nakhon Ratchasima Rajabhat University, Thailand. His research interest is in the area of data management, computer network, and data mining.