ANSF
APPLIED AND NATURAL SCIENCE FOUNDATION

Journal of Applied and Natural Science
Special Issue: 13 (SI), 135 - 141 (July, 2021)
ISSN : 0974-9411 (Print), 2231-5209 (Online)
journals.ansfoundation.org

Research Article

Trained neural network to predict paddy yield for various input parameters in Tamil Nadu, India

G. Vanitha*
Department of Computer Science, School of Post Graduate Studies, Tamil Nadu Agricultural University, Coimbatore - 641003 (Tamil Nadu), India
J. S. Kennedy
Dean, School of Post Graduate Studies, Tamil Nadu Agricultural University, Coimbatore - 641003 (Tamil Nadu), India
R. Prabhu
Department of Genetics and Plant Breeding, School of Post Graduate Studies, Tamil Nadu Agricultural University, Coimbatore - 641003 (Tamil Nadu), India
S. K. Rajkishore
Department of Environmental Sciences, School of Post Graduate Studies, Tamil Nadu Agricultural University, Coimbatore - 641003 (Tamil Nadu), India

*Corresponding author. Email: vanithag@tnau.ac.in

Article Info
https://doi.org/10.31018/jans.v13iSI.2812
Received: March 22, 2021
Revised: May 30, 2021
Accepted: June 14, 2021

How to Cite
Vanitha, G. *et al.* (2021). Trained neural network to predict paddy yield for various input parameters in Tamil Nadu, India. *Journal of Applied and Natural Science*, 13 (SI), 135 - 141. https://doi.org/10.31018/jans.v13iSI.2812

Abstract
The major objective of the present study was to explore if Artificial Neural Network (ANN) models with back propagation could efficiently predict the rice yield under various climatic conditions; ground-specific rainfall, ground-specific weather variables and historic yield data. The back propagation algorithm will calculate each expected weight using the error rate as the activity level of a unit was altered. The errors in the model during the training phase were solved during the back-propagation. The paddy yield prediction took various parameters like rainfall, soil moisture, solar radiation, expected carbon, fertilizers, pesticides, and the long-time paddy yield recorded using Artificial Neural Networks. The R^2 value on the test set was found to be 93% and it showed that the model was able to predict the paddy yield better for the given data set. The ANN model was tested with learning rates of 0.25 and 0.5. The number of hidden layers in the first layer was 50 and in the second hidden layer was 30. From this, the testing value of R square was 0.97. The observations with the ANN Model showed that i) the best result for the test set was R^2 value of 0.98, ii) the two hidden layers kept with 50 neurons in the first layer and 30 neurons in the second one, iii) the learning rate was of 0.25. With all these configurations, maximum yield is possible from the paddy crop.

Keywords: Artificial neural networks, Multilayer perceptron, Root mean square error

INTRODUCTION

Agriculture is the backbone of the Indian economy. Paddy crop rules the roost in it. The advancement in technology has led to forecasting and predicting weather conditions, resulting in an increase of yield. Various input parameters like land use, soil PH, soil moisture, fertilizers and pesticides used, the quantity of seed per hectare, *etc.,* play a vital role in predicting the crop yield. Plenty of research and study has been carried out using statistical techniques like regression model, agro-meteorological models and statistical models for the prediction of crop yield.

Nowadays, Artificial Neural Networks (ANN), a technique in data mining, has been put into use in the agricultural field for better decision making of policymakers and agricultural scientists for providing consultancy services to the farmers. Raorane and Kulkarni (2012) developed innovative approaches to predict the influence of different meteorological parameters on the crop yield using a decision tree induction approach based on long term meteorological data. Dahikar and Rode (2014) suggested the applicability of neural network technology for forecasting crop diseases. IACAT (2015) provided an outline of the work in forecasting ANN, neural network modelling and general model of the ANNs used for forecasting. According to their study, ANNs were found to be superior to the statistical models.

Kalpana *et al.* (2014) understood that statistical procedures like regression, statistical image analysis, density function and principal component analysis can be used to get these findings after studying the relevant literature on ANNs. He has explained the learning algorithm and has made a comparative analysis between statistical and neural network models in terms of terminology representations and applications. Gandhi *et al.* (2016) compared the concept of Multilayer Perceptron (MLP) with the normal statistical techniques. The errors from statistical techniques were comparatively higher than those errors from the back-propagation algorithm of MLP.

Dakshayini (2017) applied the ANN technique for predicting the severity affectation of anthracnose diseases in legume crop. Artificial Neural Networks, being self-adaptive, data driven can identify and learn correlated patterns between input data sets and corresponding target values through training. In fact, Artificial Neural Network models have been developed for paddy data for the past 20 years based on the advice of paddy experts working in the agricultural domain, and the present study takes into account such development. The paddy production depends on various input factors like the quantity of seeds sown, rainfall, soil moisture, solar radiation, expected carbon, fertilizers, pesticides, etc. Hence, crop yield prediction (Barla *et al.,* 2010; Chawla *et al.,* 2016) becomes a harder task. Here arises the reliability of Artificial Neural Networks, which can handle multi-variate nonlinear, non-parametric statistical approaches more efficiently. Artificial Neural Network models are more effective and reliable as compared to the other linear regression models for predicting the paddy yield. The main aim of the present work is to build an Artificial Neural Network (ANN) model with back propagation that could efficiently predict the rice yield under various climatic conditions; ground-specific rainfall, ground-specific weather variables and historic yield data.

MATERIALS AND METHODS

Primary data collection

The data set covered the paddy yield in Tamil Nadu district from the year 1990 to 2010 as shown in Table 1. In this implementation, sixteen input parameters and one output parameter were considered. Artificial Neural Networks (ANN) using back propagation algorithm is the most typical and widely-used model in all neural network models. It is a computational tool that acts as a biological neuron system with three layers (Li and Tian, 2003): Input layer, Hidden layer and Output layer. The input layer accepts the input data given to the model and the predicted value after computation will be produced in the output layer. The hidden layer, contains perceptron which plays a major role in transferring the input values into desired output. A certain weightage is applied to each perceptron and it is adjusted to get nearer to the desired output value. Data flow across the layers over the weighted connections. This unidirectional neural network is also known as Feed Forward Neural Network. The Artificial Network Network (ANN) was used to train and test the dataset available after pre-processing.

Back-propagation is just a way of propagating the total loss back into the neural network to know how much of the loss every node is responsible for and subsequently updating the weights to minimise the loss by giving the loss nodes with higher error rates lower weights and vice versa. (Davey, 2011 and Deshpande and Karypis, 2004).

The errors in the model during the training phase are solved during the back-propagation. The back-propagation algorithm is advantageous because the hidden units have no target values since the input units are trained using the errors of the previous layers. The training phase will continue to work until the errors in the weights are getting reduced and minimized (Lilley, 2007; Dermo, 2009 and Dubey, 2011).

Artificial neural network model development

The dataset consisting of 100 records were collected. The first step was to pre-process them by removing duplicate, unpredictable and misplaced values. This data pre-processing (Dakshayini, 2017) was again divided into training set, validation set and test set. For this dataset 75 records were reserved as training set, 15 records were occupied as validation set and the enduring 10 records were tested as test set. The training set was used to train the network until the maximum value of R^2 was grasped. The validation set was used to generalize the network. The test set was finally used to measure the performance of the network for unidentified values in the dataset.

The flow diagram (Fig.1) shows the complete steps involved in the prediction process. The input parameters included soil parameters, crop data from the initial stage. The pre-processing of data was done in order to reduce the anomalies and duplicate entries. Nearly, 75% of data were taken as Training dataset, 15% for Validation dataset and remaining 10% as Test dataset. After all the training, test and validation part was completed, finally the prediction was carried out.

The proposed algorithm involved is shown in the following steps:

Step 1: Pre-processed the data set of 100 records by removing redundant and missing values.

Step 2: Divided the data set into 75% training set (75 records), 15% as the validation set and the remaining

10% as a test set.

Step 3: Used Levenberg Marquardt algorithm for training the data set.

Step 4: Used log sig transfer function for hidden layers and purelin transfer functions for output layer.

Step 5: The feed forward back-propagation network was developed by varying the following conditions:

Number of hidden layers from 1 to 2

Number of neurons in hidden layers from 20 to 100

Learning rates as 0.25 and 0.5

Choose the network weights as random

Step 6: Repeated step 5 until the neural network model with the increased test accuracy and lower error prediction is obtained.

RESULTS AND DISCUSSION

Fig. 2 shows the statistic value R^2, used as the measure of accuracy, which was calculated using Equation 1 (Kalpana, R., 2014):

R^2 = 1 – (n-1/n-p) (SSE/SST) ………. (1)

where,

SSE is the sum of squared error, SSR was the sum of squared regression, SST was the sum squared total, n was the number of observations and p was the number of regression coefficients. The error was computed using Equation 2 (Kalpana, R., 2014):

Error = (|A – B|/|A|) x 100 ….. (2)

where A is the actual yield and B is the predicted yield obtained from the prediction model. The lower the val-

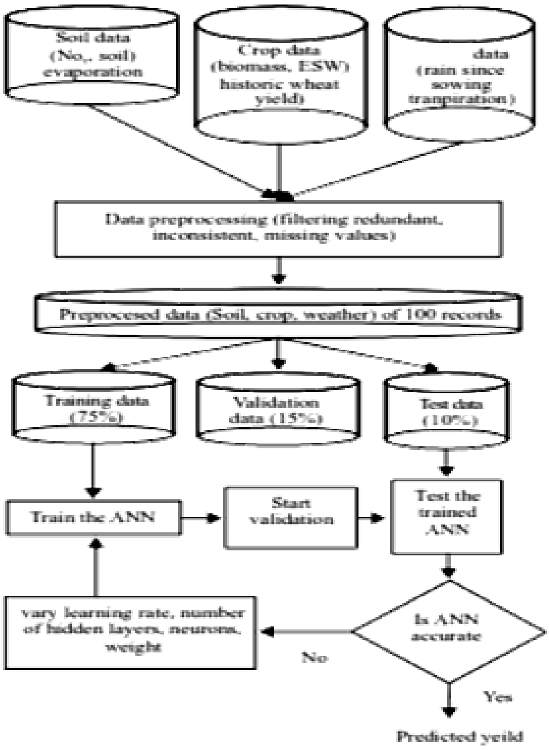
Soil data (No,. soil) evaporation
Crop data (biomass, ESW) historic wheat yield)
data (rain since sowing tranpiration)
Data preprocessing (filtering redundant, inconsistent, missing values)
Preprocessed data (Soil, crop, weather) of 100 records
Training data (75%)
Validation data (15%)
Test data (10%)
Train the ANN
Start validation
Test the trained ANN
vary learning rate, number of hidden layers, neurons, weight
Is ANN accurate
No
Yes
Predicted yeild

Fig. 1. *Flow diagram of the ANN model.*

Table 1. Sample paddy data of Thanjavur district.

Seed/ ha in kg	DAP/ ha	Urea/ ha	Potash/ ha	Soil Moisture	Soil Temperature	Rainfall	Solar Radiation	SAND_PER	SILT_PER	CLAY_PER	PH	EC	OC_PER	EX_CA	EX_MG	EX_K	ESP	Yield
0.76	0.59	0.344	0.315	0.00	0.474	0.385	0.61	0.82	0.207	0.221	0.698	0.058	0.625	0.056	0.000	0.091	0.259	0.083
0.75	0.56	0.699	0.279	0.00	0.519	0.115	1.00	0.08	0.469	0.952	0.569	1.000	0.177	0.429	0.083	1.000	0.195	0.104
0.75	0.58	0.962	0.950	0.00	0.593	0.182	0.72	0.35	0.357	0.711	0.723	0.085	0.490	0.393	0.717	0.305	0.000	0.131
0.66	0.00	0.926	0.962	0.00	0.520	0.286	0.56	0.20	0.300	0.977	0.823	0.150	0.271	0.500	0.028	0.162	0.089	0.093
0.57	0.56	0.288	0.962	0.00	0.589	0.175	0.54	0.00	0.573	0.945	0.399	0.044	1.000	0.714	0.481	0.000	0.058	0.147
0.70	0.86	0.895	0.357	0.01	0.456	0.311	0.57	0.20	0.300	0.977	0.823	0.150	0.271	0.500	0.028	0.162	0.089	0.120
0.67	0.55	0.894	0.240	0.00	0.570	0.246	0.55	0.22	0.413	0.821	0.998	0.312	0.354	0.436	0.238	0.467	0.851	0.069
0.67	0.55	0.267	0.274	0.14	0.220	0.477	0.61	0.00	0.573	0.945	0.399	0.044	1.000	0.714	0.481	0.000	0.058	0.198
0.78	0.60	0.910	0.279	0.06	0.000	0.591	0.64	0.19	0.454	0.819	0.411	0.035	0.208	0.580	0.432	0.010	0.202	0.149
0.74	0.55	0.897	0.275	0.00	0.427	0.064	0.51	0.00	0.573	0.945	0.399	0.044	1.000	0.714	0.481	0.000	0.058	0.137

Table 2. ANN results for 2 hidden layers and LR = 0.25.

ANN results for 2 hidden layers and LR = 0.25				
No. of neurons in 1st layer	**No. of neurons in 2nd layer**	**Training (R^2)**	**Validation (R^2)**	**Testing (R^2)**
20	20	0.99	0.69	0.62
	30	0.99	0.69	0.80
	40	0.1	0.83	0.95
	50	0.99	0.86	0.80
	60	0.1	0.67	0.43
	70	0.1	0.83	0.55
	80	0.89	0.72	0.38
	90	0.89	0.95	0.65
	100	0.1	0.36	0.46
30	20	0.99	0.64	0.89
	30	0.97	0.52	0.79
	40	0.1	0.02	0.64
	50	0.99	0.48	0.79
	60	0.89	0.66	0.89
	70	0.94	0.47	0.58
	80	0.1	0.90	0.56
	90	0.93	0.72	0.50
	100	0.99	0.90	0.56
40	20	0.99	0.90	0.56
	30	0.1	0.79	0.85
	40	0.99	0.65	0.85
	50	0.99	0.87	0.95
	60	0.1	0.90	0.56
	70	0.99	0.72	0.50
	80	0.87	0.90	0.56
	90	0.1	0.65	0.85
	100	0.99	0.87	0.95
50	20	0.99	0.99	0.90
	30	0.91	0.65	0.56
	40	0.1	0.52	0.79
	50	0.59	0.02	0.64
	60	0.1	0.48	0.79
	70	0.98	0.66	0.89
	80	0.99	0.47	0.58
	90	0.99	0.56	0.82
	100	0.99	0.14	0.17

Table 3. ANN results for 2 hidden layers and LR = 0.5.

ANN results for 2 hidden layers and LR = 0.5				
No. of neurons in 1st layer	**No. of neurons in 2nd layer**	**Training (R^2)**	**Validation (R^2)**	**Testing (R^2)**
20	20	0.86	0.95	0.76
	30	0.94	0.79	0.93
	40	0.83	0.68	0.79
	50	0.1	0.87	0.75
	60	0.99	0.82	0.56
	70	0.1	0.35	0.84
	80	0.1	0.32	0.46
	90	0.99	0.82	0.60
	100	0.99	0.08	0.45
30	20	0.95	0.82	0.77
	30	0.79	0.23	0.18
	40	0.98	0.83	0.80
	50	0.99	0.92	0.80
	60	0.79	0.85	0.81
	70	0.1	0.73	0.65
	80	0.73	0.69	0.49
	90	0.1	0.85	0.79
	100	0.98	0.73	0.56
40	20	0.99	0.42	0.84
	30	0.99	0.74	0.76
	40	0.99	0.75	0.94
	50	0.86	0.95	0.88
	60	0.1	0.76	0.59
	70	0.89	0.82	0.60
	80	0.99	0.08	0.45
	90	0.95	0.82	0.77
	100	0.79	0.23	0.18
50	20	0.99	0.14	0.85
	30	0.97	0.90	0.65
	40	0.97	0.17	0.64
	50	0.99	0.10	0.16
	60	0.83	0.68	0.79
	70	0.1	0.87	0.75
	80	0.99	0.82	0.56
	90	0.76	0.95	0.38
	100	0.97	0.43	0.91

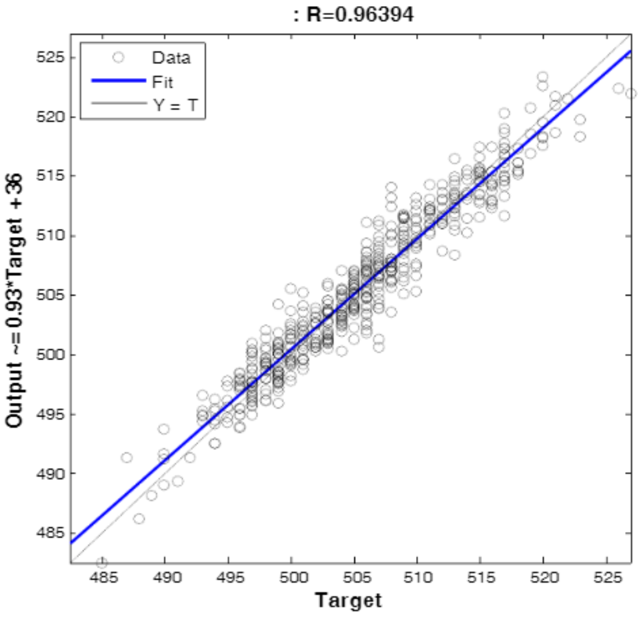

Fig. 2. *Output of the model.*

ue of error, the lesser is the predictive accuracy of the model.

Table 2 shows the results for two hidden layers with Learning Rate (LR) = 0.25. The number of neurons in the first hidden layer was kept fixed as 30 for the first time. Then, the neurons in the second hidden layer were kept varying from twenty to a hundred. This model was repeated for 20, 30, 40 and 50 neurons in the first hidden layer and varied the number of neurons from 20-100 in the second hidden layer.

The results in Table 3 show the result for two hidden layers with a learning rate LR = 0.5. The best result was obtained when the number of neurons in the first hidden layer was fixed as 20 and the second hidden layer was fixed as 30 with a testing R^2 statistic value of 0.93.

The number of neurons varied from 20-100 within each hidden layer. The Artificial Neural Network model was tested with learning rates of 0.25 and 0.5. The best result was obtained when the number of hidden layers in the first hidden layer was set as 50 and the second hidden layer was 20 with a testing R^2 statistic value of 0.96.

Finally, it was observed that the ANN model with the following configurations gave the best result of 0.97 (R^2 statistic value) for the test set:

Two hidden layers with 50 number of neurons in the first layer and 20 number of neurons in the second layer were the best value fixed.

Learning rate with 0.25 value gave the optimised result.

The back-propagation ANN model used is more advantageous than the other forecasting models since the hidden units have no target values and the error rate is very low. But the time series analysis model does not give a precise outcome (Mariappan and Austin, 2017). But, in nonlinear FFBN and linear PLSR models for rice prediction, the climate data was omitted and hence the reliability and accuracy of the model is a major drawback of this model. (Hossain et al., 2017) . In another model using Support Vector machine, the process was based on image analysis results that are not accurate as soil conditions are not considered (M.Shashi 2019). To overcome all the above models, back-propagation algorithm using ANN model gives the best result.

In this work, paddy yield prediction had taken into account all the parameters like rainfall, soil moisture, solar radiation, expected carbon, fertilizers, pesticides and the long-time paddy yield records using Artificial Neural Networks. The R^2 value on the test set was found to be 93% and it showed that the model was able to predict the paddy yield better for the given data set. The future work may be extended to other crops in various districts based on the same model.

Conclusion

The prediction of crop yield plays a vital role in the agricultural field. In this regard, paddy was taken into account and the yield was predicted based on the input parameters like climate, soil nutrients, fertilizers, pesticides, seed varieties, etc. The result found the R^2 value of 93% with the model developed. The outcome of this research work may help the agricultural officers to predict crop conditions for improving paddy yield. In future, a generalized prediction model for various crops by taking into account various parameters can be developed to reach the ultimate object for the maximum yield of every crop with no negative influences on it.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. Barla, M., Bielikova, M., Ezzeddinne, A.B., Kramar, T., Simko, M. & Vozar, O. (2010). On the impact of adaptive test question selection for learning efficiency. *Computers & Education,* 55(2), 846-857.
2. Chawla, V., Naik, H.S. & Akintayo, A. (2016). A bayesian network approach to county-level corn yield prediction using historical data and expert knowledge. *Proceedings of the 22nd ACM SIGKDD Workshop on Data Science for Food, Energy and Water*, San Francisco, CA, USA, 2016.
3. Dahikar, S. & Rode, S. (2014). Agricultural crop yield prediction using artificial neural network approach. *International Journal of Innovative Research in Electrical, Electronic Instrumentation and Control Engineering,* 2(1), 683-686.
4. Dakshayini, P. (2017). Rice crop yield prediction using Data mining techniques: An overview. *International Journal of Advanced Research in Computer Science and Soft-*

ware Engineering, 7, 427-431.

5. Davey, T. (2011). Educational testing service. In: A guide to computer adaptive testing systems, Technical issues in large-scale assessment (TILSA), State Collaborative on Assessment and Student Standards (SCASS), 2011.
6. Dermo, J. (2009). E-Assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology,* 40(2), 203-214.
7. Deshpande, M. & Karypis, G. (2004). Selective Markov models for predicting web page accesses. *ACM Transactions on Internet Technology,* 4, 163-184 ,2004.
8. Dubey, M. (2011). Effective e-learning Design, Development and Delivery', University Press, 7.
9. Gandhi, N., J. Leisa, Armstrong & O. Petkar (2016). Predicting rice crop yield using bayesian networks. *Conference on Advances in Computing, Communications and Informatics,* Jaipur, India, 2016
10. Hossain, M.A., Uddin, M.N., Hossain, M.A. & Jang, Y.M. (2017). Predicting rice yield for Bangladesh by exploiting weather conditions, Jeju, South Korea, pp. 1–6
11. Kalpana, R., Shanti, N. & Arumugam, S. (2014). A survey on data mining techniques in agriculture. *International Journal of Advances in Computer Science and Technology,* 3(8), 426-431.
12. Li, Z. & Tian, J. (2003). Testing the suitability of Markov chains as web usage models. *Proceedings of 27th annual International Computer Software and Applications Conference,* Dallas, TX, USA, 356-361.
13. Lilley, M. (2007). The development and application of computer adaptive testing in a higher education environment, *Ph.D. Thesis,* University of Hertfordshire, UK ,2007.
14. Mariappan, A.K., Austin Ben & Das, J. (2017). A paradigm for rice yield prediction In Tamil Nadu. In: International Conference on Technological Innovations in ICT for Agriculture and Rural Development, Chennai, India, pp. 1–4.
15. M. Shashi, Y. (2019). Atmospheric temperature prediction using support vector machines. *International Journal of Computer Theory and Engineering*, 1 (1), 55-58.
16. Raorane, A.A. & R.V. Kulkarni. (2012). Data mining-An effective tool for yield estimation in the agricultural sector. *International Journal of Engineering Trends and Technology,* 1(2), 75-79.