

Original Research Article

<https://doi.org/10.20546/ijcmas.2021.1001.007>

Prediction of Paddy Yield in Thiruvarur District using Artificial Neural Network

G. Vanitha^{1*} and B. Sivasankari²

¹Tamil Nadu Agricultural University, Coimbatore, India

²Agriculture College and Research Institute, Tamil Nadu Agricultural University, Madurai, India

*Corresponding author

ABSTRACT

Agriculture is the major deciding factor of Indian economy and Paddy is the principal crop cultivated extensively in all the districts of Tamil Nadu. Thiruvarur district in Tamil Nadu leads paddy cultivation with an area of 1,78,080 ha in crop yield production. Eventually, Artificial Neural Network (ANN) techniques have emerged to be important for predicting and maximising the crop yield for the benefit of farmers. This research is based on the development of trained Neural Network models for predicting the paddy yield by varying the input parameters including both controllable and uncontrollable factors. The models have been experimented with different input parameters of paddy and training patterns. For this purpose, real data set from the Department of Agricultural Meteorology, Department of Soil Sciences, Department of Economics, Directorate of CARDS, Tamil Nadu Agricultural University, Coimbatore and Tamil Nadu Rice Research Institute, Aduthurai, were collected. The collected data were intensively studied, the modus operandi using the data was arrived at, and taken for various experiments. The experiments show that the trained neural network produced a minimum error which indicated that the test model is capable of predicting and maximising the paddy yield in Thiruvarur. The major objectives of this paper are to: (i) explore if Artificial Neural Network models with back propagation could efficiently predict rice yield in Thiruvarur district under various climatic conditions; ground-specific rainfall, ground-specific weather variables (sunshine hours, solar radiation, maximum and minimum temperature, daily wind speed values) and historic yield data; (ii) analyse the changes of model performance with variations of ANN model parameters; and (iii) calculate the accuracy by which crop yield prediction is made.

Keywords

Data Mining, Back propagation feed forward algorithm, R squared value, Yield Prediction, Forecasting, Model generation

Article Info

Accepted:
04 December 2020
Available Online:
10 January 2021

Introduction

Agriculture is the backbone of Indian economy, Paddy crop rules the roost in it. The advancement in technology has led to forecasting and predicting weather conditions

which results in the increase of yield. Various input parameters like land use, soil PH, soil moisture, fertilizers and pesticides used, 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 prediction of crop yield.

Nowadays, Artificial Neural Networks (ANN), a technique in data mining has been put into use in agricultural field for better decision making of policy makers and agricultural scientists for providing consultancy services to the farmers. A.A. Raorane *et al.*, 2011 developed innovative approaches to predict the influence of different meteorological parameters on the crop yield using decision tree induction approach based on long term meteorological data. Dewolf and Francl suggested the applicability of neural network technology for forecasting the crop diseases. Zhang *et al.*, 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.

Titterington [3] 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 given an explanation on the learning algorithm and has made a comparative analysis between the statistical regression and neural network models in terms of terminology representations, and applications.

Gaudart *et al.*, [4] compared the performance of Multilayer Perceptron (MLP) and that of linear regression for epidemiological data with respect to quality of prediction and robustness to deviation from underlying assumptions of homoscedasticity normality, and independence of errors and it was found that MLP performed better than the linear regression model. Chakraborty *et al.*, [5] 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 production of paddy depends on various input factors like quantity of seeds sown, rainfall, soil moisture, solar radiation, expected carbon, fertilizers, pesticides, etc. Hence, the prediction of crop yield becomes harder task and here arises the reliability of Artificial Neural Networks which can handle multi-variate non-linear, 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.

Materials and Methods

Primary data collection

The data set covers the paddy yield in Thiruvavur district from the year 1990 to 2010 as shown in the table 1. In this implementation, sixteen input parameters and one output parameter has been 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 which acts like a biological neuron system with three kinds of layer [6]: 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 (Fig. 1 and 2).

The back propagation algorithm is a method of computing the error derivative (EW) in ANN. The back propagation algorithm computes each EW by computing the rate at which the error changes as the activity level of a unit is changed (Hinton, 1992). Basically, the errors are back propagated from the output units towards the input unit during training phase. This algorithm is important as the hidden units do not have target values and these units should be trained based on errors from the previous layers. The weight value continuously gets update as the errors are back-propagated. Training phase will continue until the errors in the weights are minimized (Rojas, 1996 and Hassoun, 1995).

Artificial Neural Network model development

The dataset of 100 records were pre-processed by removing redundant, inconsistent and missing values. This pre-processed data [14] was again divided into training set (75%) with 75 records, 15% as validation set with 15 records and the remaining 10% as test-set with 10 records. The training set was used to train the network until the maximum value of R^2 was reached. The validation set was used to generalize the network. The test set was used to measure the performance of the network for unknown values.

The flow diagram as shown below Fig. 1 shows the complete steps involved in the prediction process. The input parameters include 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-process the data set of 100 records by removing redundant and missing values.

Step 2: Divide the data set into 75% training set (75 records), 15% as validation set and remaining 10% as test set.

Step 3: Use Levenberg Marquardt algorithm for training the data set

Step 4: Use logsig transfer function for hidden layers and purelin transfer functions for output layer.

Step 5: The feed forward back-propagation network is 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: Repeat step 5 until the neural network model with increased test accuracy and lower error prediction is obtained.

Results and Discussion

Figure 3 shows the statistic value R^2 , used as the measure of accuracy which is calculated using the Equation 1:

$$R^2 = 1 - (n-1/n-p) (SSE/SST) \quad - (1)$$

where, SSE is the sum of squared error, SSR is the sum of squared regression, SST is the sum squared total, n is the number of observations and p is the number of regression coefficients. The error is computed using the Equation 2:

$$\text{Error} = (|A - B|/|A|) \times 100 \quad - (2)$$

Where, A is the actual yield and B is the predicted yield obtained from the prediction model. The lower the value of error, the lesser is the predictive accuracy of the model.

Table 2 shows the results for 2 hidden layers with Learning Rate (LR) = 0.25. At first, the number of neurons in first hidden layer was kept fixed (20), then the neurons in second hidden layer was varied from 20-100. This experiment was repeated for 20, 30, 40 and 50 neurons in the first hidden layer and varied the number of neurons in the second hidden layer from 20-100 (Fig. 4 and 5).

Fig.1 Flow diagram of the model

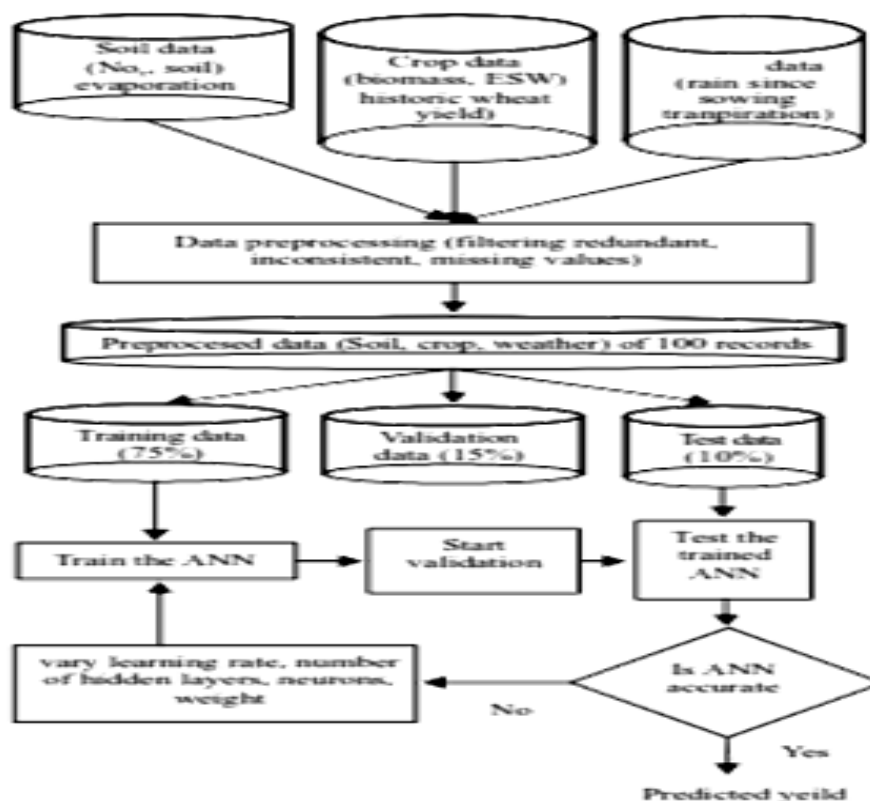


Table.2 ANN results for 2 hidden layers and LR = 0.25

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	0.99	0.95	0.65
	100	0.1	0.36	0.46
	30	20	0.99	0.64
30		0.97	0.52	0.79
40		0.1	0.016	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	0.59
90		0.93	0.72	0.50
100		0.99	0.90	0.56
40		20	0.99	0.86
	30	0.1	0.79	0.85
	40	0.99	0.79	0.70
	50	0.99	0.26	0.88
	60	0.1	0.53	0.88
	70	0.99	0.87	0.36
	80	0.87	0.77	0.50
	90	0.1	0.86	0.80
	100	0.99	0.89	0.95
	50	20	0.99	0.90
30		0.91	0.79	0.24
40		0.1	0.65	0.74
50		0.59	0.75	0.63
60		0.1	0.58	0.74
70		0.98	0.89	0.65
80		0.99	0.56	0.82
90		0.99	0.14	0.17
100		0.89	0.58	0.11

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
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	0.66	0.62
80		0.73	0.69	0.49
90		0.1	0.73	0.65
100		0.98	0.06	0.81
40		20	0.99	0.42
	30	0.99	0.74	0.76
	40	0.86	0.71	0.76
	50	0.86	0.95	0.88
	60	0.1	0.76	0.48
	70	0.99	0.61	0.88
	80	0.1	0.77	0.70
	90	0.99	0.14	0.68
	100	0.1	0.40	0.92
	50	20	0.99	0.74
30		0.97	0.90	0.65
40		0.99	0.17	0.64
50		0.99	0.10	0.16
60		0.96	0.36	0.73
70		0.75	0.75	0.81
80		0.97	0.43	0.91

Table.1 Sample Paddy data

Seed/ha in kg	DA P/ha	Urea/ha	Potash/ha	Soil Moisture	Soil Temperature	Rainfall	Solar Radiation	SAND_PER	SILT_PER	CLAY_PER	PH	EC	OC_P ER	EX_C A	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

Fig.2 Output of the model

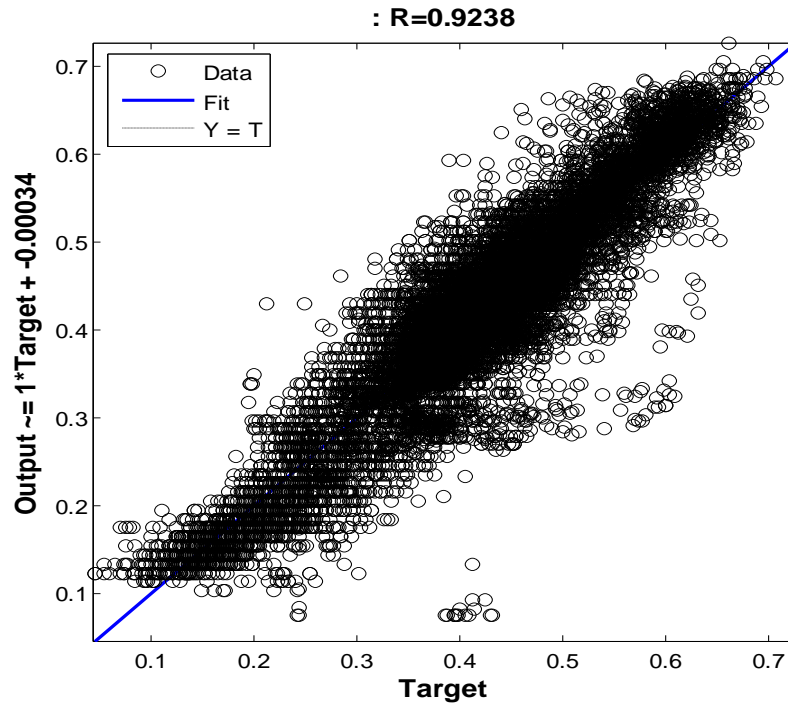


Fig.3 Training state plot of Paddy Yield Prediction system

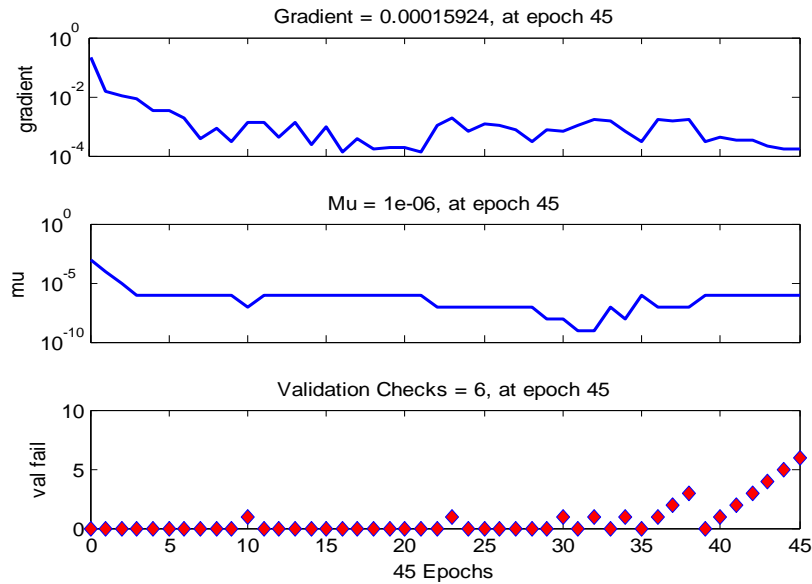


Fig.4 Training state plot of paddy yield prediction system

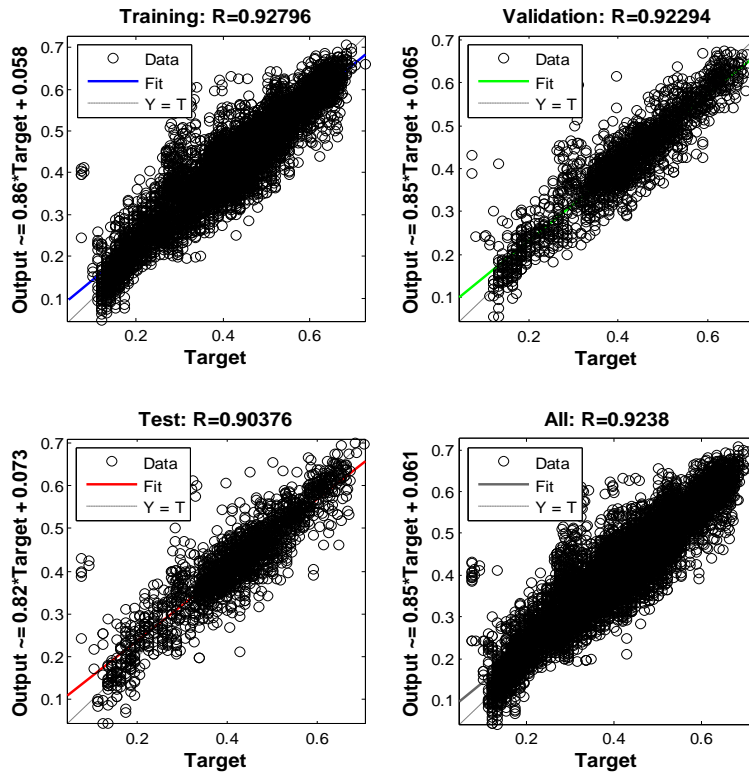
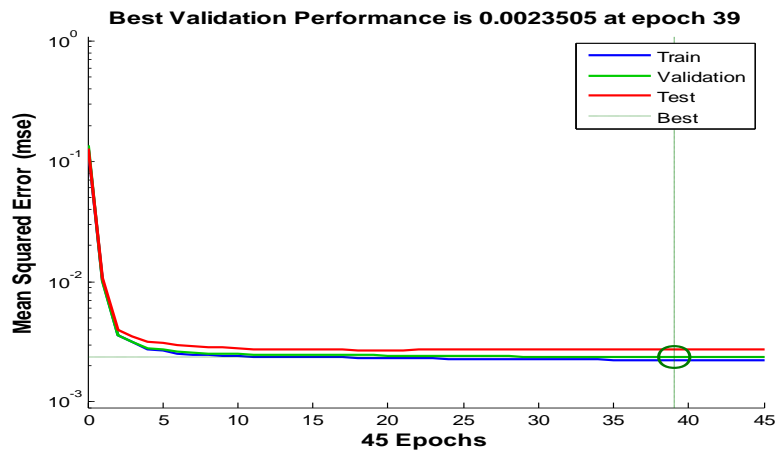


Fig.5 Performance plot of paddy yield prediction system



The table 3 shows the result for 2 hidden layers with learning rate $LR = 0.5$. The best result was obtained when the number of neurons in the first hidden layer was 20 and the second hidden layer was 30 with a testing

R^2 statistic value of 0.93. Within each hidden layer the number of neurons was varied from 20-100. The ANN 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 1st hidden layer was 50 and 2nd hidden layer was 20 with a testing R^2 statistic value of 0.93.

Finally, it is observed that the ANN model with the following configurations gave the best result of 0.93 (R^2 statistic value) for test set as shown in Table 3.

Two hidden layers with 50 number of neurons in 1st layer and 20 number of neurons in 2nd layer.

Learning rate of 0.25.

The result shows that the paddy yield prediction can be done reliably using back propagation neural network algorithm under various parameters like rainfall, soil moisture, solar radiation, expected carbon, fertilizers, pesticides and the past records of paddy yield. In conclusion the crop yield prediction is a factor in agricultural field which has to be reckoned with.

In this work, paddy yield prediction has taken into account various 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 shows 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 as well in various districts based on the same model.

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 parameter can be developed to reach the ultimate object "Maximum yield from every crop with no detrimental influences on it".

References

- 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, vol.55, no.2, pp.846-857
- Dahikar, S., and S. Rode, "Agricultural crop yield prediction using artificial neural network approach", International Journal of Innovative Research in Electrical, Electronic Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683-686, 2014
- Dakshayini Patil, 2017, "Rice crop yield prediction using Data mining techniques: An Overview, International Journal of Advanced Research in Computer Science and Software Engineering, vol.7, pp.427-431.
- Dermo, J 2009, 'E-Assessment and the student learning experience: A survey of student perceptions of e-assessment', British Journal of Educational Technology, vol.40 no.2, pp.203-214
- Deshpande, M and Karypis, G 2004, 'Selective Markov models for predicting web page accesses', ACM Transactions on Internet Technology, vol.4, pp. 163-184
- Kalpana, R., N. Shanti and Arumugam, S. "A survey on data mining techniques in Agriculture", International Journal of advances in Computer Science and Technology, vol. 3, No. 8,426- 431, 2014.
- 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, pp. 356-361
- Lilley, M 2007, 'The development and application of computer adaptive

- testing in a higher education environment’, PhD thesis, University of Hertfordshire, UK
- Madhuri Dubey, 2011, ‘Effective e-learning Design, Development and Delivery’, University Press.7
- Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, 2016, “Predicting Rice Crop Yield Using Bayesian Networks”, Conference on Advances in Computing, Communications and Informatics.
- Raorane, A.A. and R.V. Kulkarni, “Data Mining-An effective tool for yield estimation in the agricultural sector”, International Journal of Engineering Trends and Technology, vol. 1, No.2, pp. 75-79, 2012.
- Tim Davey, Educational Testing Service, ‘A Guide to Computer Adaptive Testing Systems’, Technical Issues in Large-Scale Assessment (TILSA) State Collaborative on Assessment and Student Standards (SCASS), 2011
- Vikas Chawla, Hsiang Sing Naik, Adedotun Akintayo, 2016, “A Bayesian Network approach to County-Level Corn Yield Prediction using historical data and expert knowledge”, Data Science for Food, Energy and Water.

How to cite this article:

Vanitha, G. and Sivasankari, B. 2021. Prediction of Paddy Yield in Thiruvarur District using Artificial Neural Network. *Int.J.Curr.Microbiol.App.Sci.* 10(01): 47-56.
doi: <https://doi.org/10.20546/ijcmas.2021.1001.007>