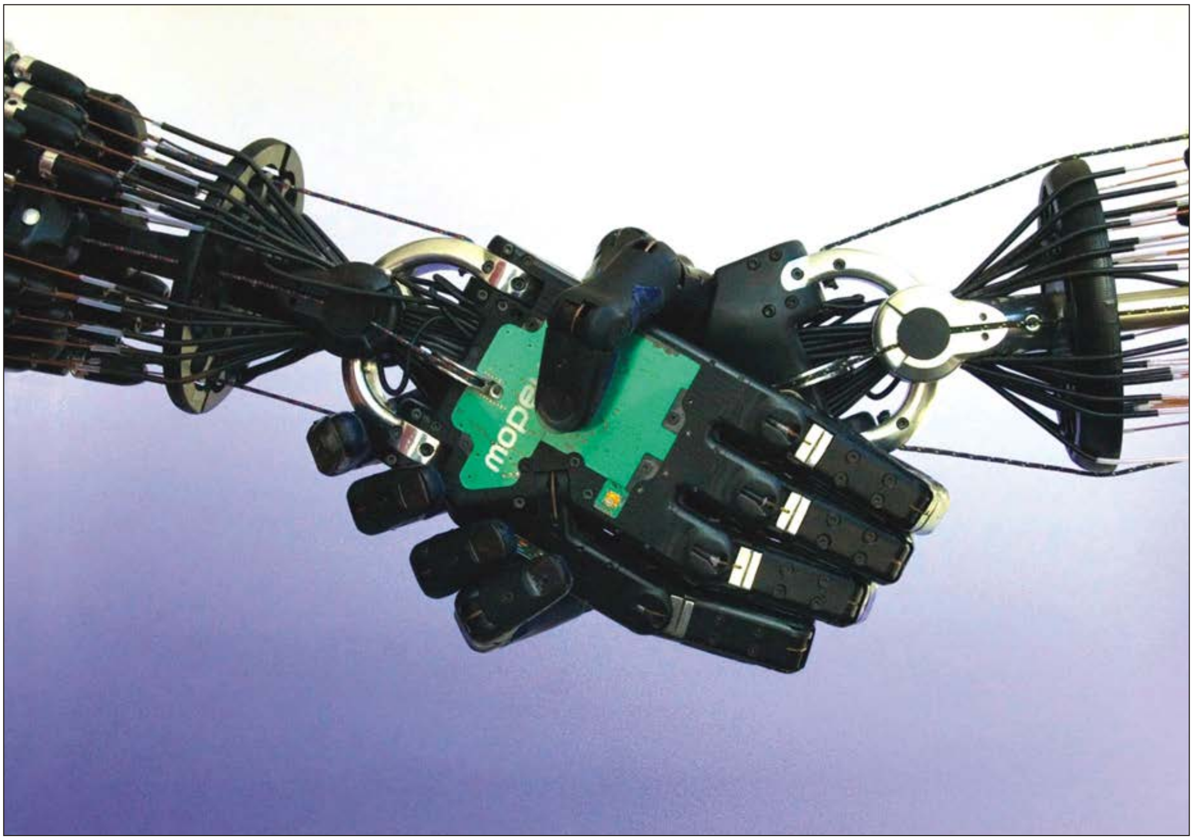

Journal of Artificial Intelligence

ISSN 1994-5450

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS Journal of Artificial Intelligence

ISSN 1994-5450
DOI: 10.3923/jai.2016.23.32

Research Article

A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction

K. Aditya Shastry, H.A. Sanjay and Abhijeeth Deshmukh

Nitte Meenakshi Institute of Technology, Govindpura, Gollahalli, Yelahanaka, Bangalore, Karanataka, 560064, India

Abstract

Background: Selection of the crop for planting is one of the major challenges faced by farmers. Crop selection is influenced by many factors like the weather, nature of soil, market, etc. Weather and soil type are the major factors which affect the crop yield. Crop yield prediction helps the farmers in the selection of the crop for plantation. Crop yield can be accurately predicted by considering the parameters like nature of the soil, amount of rain, crop characteristics, etc. **Methodology:** There are couple of methods which can be used to predict crop yield. Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) are two well-known prediction techniques. In this study prediction of the wheat crop yield is done by considering parameters like amount of rainfall, crop biomass, soil evaporation, transpiration, Extractable Soil Water (ESW) and amount of fertilizer applied (NO_3). Default-Artificial Neural Networks (D-ANN) is a ANN with only one hidden layer. In this study Customized ANN (C-ANN) is developed by varying the number of hidden layers, number of neurons in the hidden layer and the Learning Rate (LR). Experiments are conducted to compare the C-ANN with MLR and D-ANN models on the same dataset using R^2 statistic and percentage prediction error. **Results:** Results show that the C-ANN model performs better with a higher R^2 statistic and a lower percentage prediction error than the MLR and D-ANN models on the test dataset. **Conclusion:** Prediction of crop yield is very important in the community of agriculture. In this study wheat yield was predicted by considering its different parameters. Better wheat yield was predicted by using C-ANN model.

Key words: Artificial neural networks, multiple linear regression, yield, parameters, customized ANN

Received: October 07, 2015 Accepted: December 10, 2015 Published: June 15, 2016

Citation: K. Aditya Shastry, H.A. Sanjay and Abhijeeth Deshmukh, 2016. A parameter based customized artificial neural network model for crop yield prediction. J. Artif. Intel., 9: 23-32.

Corresponding Author: K. Aditya Shastry, Nitte Meenakshi Institute of Technology, Govindpura, Gollahalli, Yelahanaka, Bangalore, Karanataka, 560064, India

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

Data mining (the investigation venture of the "Learning Discovery in Databases" procedure), a field at the intersection of software engineering and measurements is the procedure that endeavors to find designs in extensive information sets. It uses techniques at the intersection of database systems, machine learning, statistics and artificial intelligence. The objective of data mining is to extract data from an information set and change it into a justifiable structure for further use.

The genuine data mining assignment is the semiautomatic or automatic analysis of large amounts of information to extract already unknown fascinating examples, like unusual records and conditions and gatherings of information records. These examples can then be seen as a sort of input data summary and may be utilized as a part of further examination or, for instance, in predictive analysis and machine learning. For instance, the data mining step may recognize different gatherings in the information, which can then be utilized to get more precise forecast results with the help of predictive modeling.

From the past few years predictive modeling is practiced in only few areas in this competitive world. The big data phenomenon is a tool which is used for data analysis in new applications in order to increase the adoption of predictive models.

The way predictive models produce value is simple in concept; they make it possible to make more right decisions, more quickly and with less expense. They can provide support for human decisions, making them more efficient and effective, or in some cases, they can be used to automate an entire decision-making process. The motivation behind the proposed work is to build a predictive model which can be used for predicting the crop yields by providing different attributes on which crop yield is dependent.

Agriculture is a business with risk and reliable crop yield prediction is vital for decisions related to agriculture risk management[1]. Yield prediction of crops like wheat, corn and rice is important in economic programming in the global scene. It is also important in giving suggestions to the farmers regarding particular crops. An accurate estimation of crop yield also helps agriculture agencies in planning supply chain decisions like production scheduling[1].

Predictive analytics are used to determine the probable future outcome of an event. These tools incorporate more sophisticated analytical strategies and include data mining and modeling[2]. It has been successfully used in climatic predictions, power predictions, etc. Most of the times, prediction is done by considering historical data about the entity being predicted. Various prediction models like Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), natural heuristics, etc. are available which can be used for prediction.

In this study, a dataset of 50 records is obtained from Agriculture Production Simulator (APSim)[3] version 7.5. The data set contains various weather, crop, soil and historic yield information. This dataset is preprocessed for removal of outliers, redundant and missing values. The data set is then divided into training set (70%) of 35 records, validation set (15%) of 7 records and testing set (15%) of 8 records.

Customized-Artificial Neural Networks (C-ANN) is developed by varying the number of hidden layers, number of neurons in hidden layer and the Learning Rate (LR). This C-ANN is applied on the dataset to predict yields of wheat. Same data set is used in MLR and D-ANN to predict crop yield and the results are compared.

Several types of prediction models like artificial neural networks, multiple linear regression and support vector regression and Adaptive Neuro Fuzzy Inference System (ANFIS) applied on different crops like wheat and tomato for crop yield prediction are investigated.

Ruβ[4] discusses regression techniques like neural networks Multi-Layer-Perception (MLP), regression tree, support vector regression on selected agriculture data. MLP has been used to predict wheat yield from fertilizer and additional sensor input. They conclude that support vector regression can serve as a better reference model for yield prediction. Ramesh and Vardhan[5] presented an analysis of past climatic variations and its impact on agricultural production to get the climate variability on agriculture. De Leon and Jalao[6] investigated the development of a crop prediction model framework by considering 13 agronomic variables.

Ghodsi *et al.*[7] presented an analysis of the effects of climate factors on wheat yield. Experiment was performed in Iran country to obtain models suitable for yield estimation and regional grain production prediction. They compare real wheat production with ANN output in the last five years and show that the ANN model is a suitable way of predicting wheat production. Veenadhari *et al.*[8] review the research studies on application of data mining techniques in agricultural field. Ruβ *et al.*[9] discuss how neural networks can be used for predicting the yield of wheat from cheaply available data obtained in the years 2003 and 2004 in Germany.

Paswan and Begum[10] provided a comprehensive review of literature comparing feed-forward neural networks and traditional statistical methods, viz., linear regression with respect to prediction of agricultural crop production. Qaddoum *et al.*[11] proposed an automatic tomato yield predictor to assist the human operators in anticipating more effectively weekly fluctuations and avoid problems of both over demand and overproduction if the yield cannot be predicted accurately.

Singh and Prajneshu[12] have described a particular type of "Artificial Neural Network (ANN)", viz., Multilayered Feed Forward Artificial Neural Network (MLFANN). Parekh and Suryanarayana[13] have carried out a study to determine the predominance of various meteorological data on yield of wheat.

Most of the works have made use of linear regression models for prediction of crop yield. But the yield of a crop has a non-linear relationship with input parameters like weather, soil and crop. Hence non-linear models like ANN are better suited for predicting crop yield. In most of the works, critical parameters like rainfall, soil evaporation, transpiration, biomass, Extractable Soil Water (ESW) and amount of fertilizer (NO_3) are not considered for crop yield prediction. In this study, customization of artificial neural networks is done to predict the wheat yield by considering these critical parameters along with historic wheat yield information.

MATERIALS AND METHODS

This study describes artificial neural networks and multiple linear regression which are two well known prediction techniques. The ANN is a non-linear model which predicts better when there is a non-linear relationship between the input parameters. The MLR is a linear model which predicts better when there is a linear relationship between the input parameters.

Artificial neural networks: Neural network models in artificial intelligence are usually considered as artificial neural networks. These are basically simple mathematical models associated with particular learning algorithm. The common use of ANN model means the definition of a class of functions. The members of the class are obtained by varying parameters connection weights or specifying the structure of the architecture by changing number of neurons and their connectivity.

Figure 1 shows the network consists of three layers:

- **First layer:** It has input neuron which accepts the data and sends it to second layer of neurons

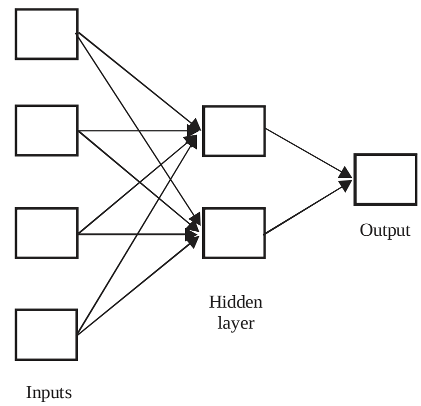

Fig. 1: Artificial neural network

- **Second layer:** It can also be called as hidden layer which accepts the data from input layer processes it and sends it to third layer of output neurons
- **Third layer:** It accepts processed data from hidden layer and creates output

More complex networks may have more number of hidden layers so as to process complex data and produce required output. They use parameters called "Weights" that manipulate the data in the calculation. An ANN is typically defined by three types of parameters:

- The interconnection pattern between different layers
- The selected learning process used to update weights of the interconnections
- The activation function that converts inserted weighted input to its output activation

Figure 2 depicts that mathematically, a neurons network function can be defined as f $(\sum w_i, x_i)$.

The x1, x2 ⋯xn are the inputs, w1, w2 ⋯ wn are the weights, which in combination forms the output function of the architecture. This can be represented as a mathematical function with arrows explaining the dependencies between variables[14] as in Eq. 1:

$$y = \theta \sum_{j=1}^{n} w_i x_i - \mu \quad (1)$$

where, θ(•) is a unit step function, w_i are the weights associated with the ith input.

Multiple linear regression: In statistics, linear regression is an approach for modeling the relationship between a dependent variable y and one or more explanatory variables denoted

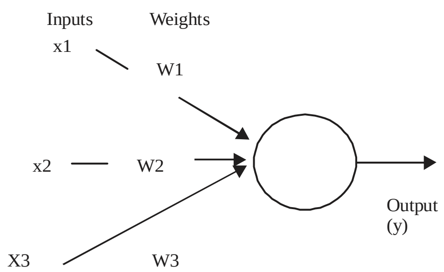

Fig. 2: Structure of a neuron

by x. The case in which only one explanatory variable is used is called simple linear regression. For more than one explanatory variable the process is called multiple linear regression. The general equation for multiple linear regression[15] is shown in Eq. 2:

$$y = \beta_0 + \beta_1 x_1 + \beta_1 x_1 + \cdots . + \beta_n x_n \tag{2}$$

where, β_0, β_1, $\cdots$ β_n are coefficients to be estimated, x_1, x_2, $\cdots$ x_n are explanatory variables (inputs), y is the dependent variable (output).

In regression, given data are modeled using linear predictor functions and unknown parameters are estimated from the data. Linear models was first type of regression analysis to be studied vastly and to be used more often in practical examples, the reason behind it is linear relations are easy to depict, because of it linear nature. Fitting of the nonlinear variables is more difficult to study and elaborate[15].

Data source and description: This study describes the source of data from which wheat data set was obtained to perform our experiments. It also describes the training data set, validation data set and test data set.

The wheat data set was obtained from Agricultural Production System Simulator (APSim) version 7.5. APSim collected the data as part of (ACIAR) Australian Centre for International Agricultural Research) project "Expanding Rabi Season Cropping in Southern Bangladesh" in 2012.

The dataset consists of 50 records from 16-03-1960 to 17-03-2009 from North Joynags, Bhola region, Southern Bangladesh with AP soil number: 673. The soil texture was silt with latitude and longitude of 22.585 and 90.666, respectively.

The data set is divided into 70% as training set (35 records), 15% as validation set (7 records) and the other

Table 1: Sample wheat data set[2]

Bio-mass	ESW	NO_3	Rain	Trans-piration	Soil-Evapo ration	Wheat yield
5253.3	105.3	28.3	33	140.148	65.963	2006.1
6905.3	113.5	25.5	106	171.259	88.736	2908.5
5911.9	94.2	23.8	39	145.226	75.363	2135.4
5163.9	113.2	30.9	34	139.884	58.763	1819.2
5739.7	98.5	25.8	8	132.416	56.592	2087.2
5822.7	93.5	24.1	35	148.363	71.335	2572.1
5163.3	98.4	27.3	32	133.102	63.143	1952.5

ESW: Extractable soil water

15% as test set (8 records) for experimentation purpose. The training set is presented to the D-ANN and C-ANN during training and the network is adjusted according to its error. The validation set is used to measure the network generalization and to halt training when generalization stops improving. The testing set has no effect on training and provides an independent measure during and after training.

Table 1 depicts a sample of wheat data set used for our experiments. The biomass is the accumulated energy in plants, ESW is extractable soil water, NO_3 is the nitrogen content present in soil, transpiration is the amount of water evaporated from the leaf, soil evaporation is the amount of water evaporated from soil and wheat yield is the historic yield from 1960-2009.

Parameters for wheat yield prediction: This study describes the parameters that were used for predicting the wheat yield. The parameters were selected based on the experiments conducted by Dalgliesh *et al*.[3].

The description of the parameters considered by us for wheat yield prediction is given below:

- Biomass (kg ha^{-1}) is the accumulated energy in plants
- ESW (mm) is the extractable soil water
- NO_3 (kg ha^{-1}) is nitrogen content present in soil
- Rain (mm) is amount of rainfall since sowing
- Transpiration (mm) is the amount of water evaporated from the leaf
- Soil evaporation (mm) is the amount of water evaporated from soil
- Historic wheat yield (kg ha^{-1}) from 1960-2009

Default artificial neural network (D-ANN): The D-ANN is a two layer feed-forward network with sigmoid hidden neurons and linear output neurons. It consists of one input layer, one hidden layer and one output layer. The network was trained using the Levenberg-Marquardt backpropogation algorithm as shown in Table 2.

Table 2: Algorithm for default ANN[14]

Input: Experimental data set of weather data, crop data and soil data
Output: Predicted wheat yield for the experimental dataset are as follows:

- Initialize the weights to small random values
- Randomly choose an input pattern $x^{(\mu)}$
- Propagate the signal forward through the network
- Compute ∂_i^L in the output layer ($O_i = y_i^L$)

$$\delta_i^L = g'(h_i^L)[d_i^\mu - y_i^L]$$

- Compute the deltas for the preceding layers by propagating error backwards:

$$\delta_i^l = g'(h_i^L)\sum_j w_{ij}^{l+1}\delta_j^{l+1} \quad \text{For } l = (L-1) \ldots 1$$

- Update the weights using

$$\Delta w_{ji}^l = \eta\delta_i^l y_j^{l-1}$$

- Go to step 2 and repeat the next pattern until the error in the output layer is below a pre-specified threshold and the maximum number of iterations is reached

In this study, the D-ANN is trained using the training data set. The number of neurons in the hidden layer was varied from 20-100. Two learning rates 0.25 and 0.5 were considered. The experiments were repeated for each of the learning rates with different number of neurons in the hidden layer. The D-ANN which gave the best R^2 value for the testing set and the lowest percentage prediction error was selected.

Customized artificial neural network (C-ANN): In this study customization of ANN is described to generate the C-ANN model. This is done by varying the number of hidden layers, the number of neurons in each hidden layer and the learning rate.

The algorithm for the C-ANN used for predicting wheat yield is shown in Table 3.

The number of hidden layers considered was 2. For each hidden layer, the number of neurons in the layer was varied from 20-100. The configuration of C-ANN which provided the best R^2 value and the lowest percentage prediction error for the testing set was selected.

R-Square (R^2) statistic was used as a measure of accuracy. The ANN model whose test set had the highest R-square value was selected as the best ANN model. It had the following configurations:

- Two hidden layers with 50 neurons in first layer and 20 neurons in second layer using logsig transfer function
- Learning rate = 0.25
- Output layer using purelin transfer function

Table 3: Algorithm for customizing ANN

Input: Experimental data set of weather data, crop data and soil data
Output: Predicted wheat yield for the experimental dataset are as follows:

- Preprocess the data set of 50 records by removing redundant, missing and inconsistent values
- Split the data set into 70% training set (35 records), 15% as validation set (7 records) and other 15% as test set (8 records)
- Use Levenberg Marquardt algorithm for training
- Use logsig transfer function for hidden layers and purelin transfer function for the output layer
- Customize feed forward back-propagation network by varying the following parameters:
  - Number of hidden layers (1-2)
  - Number of neurons in hidden layers (20-100)
  - Learning rates (0.25, 0.5)
  - network weights (random)
- Repeat step 5 until ANN model with highest test accuracy and lowest percentage predicton error is obtained

Figure 3 shows the methodology followed for customizing the artificial neural network. The data set of 50 records was pre-processed by removing inconsistent, redundant and missing values. This pre-processed data was then split into 70% as training set (35 records), 15% as validation set (7 records) and the remaining 15% as test set (8 records). The training set was used to train the C-ANN until the maximum R^2 value 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. This test set was used as a final measure of accuracy for comparing the performance of the MLR, D-ANN and C-ANN models.

Multiple linear regression: Multiple Linear regression was performed using Matlab R2013a. The following MLR equation was applied on the training data set of wheat as in Eq. 3:

$$y = -1901.8+0.6x_1+2.6x_2-15.6x_3-0.4x_4+5.3x_5+2.3x_6 \qquad (3)$$

where, y is predicted yield of wheat crop, x_1 is biomass (kg ha^{-1}), x_2 is ESW (mm), $x_,$ is NO_3 (kg ha^{-1}), x_4 is rain since sowing (mm), x_5 is transpiration (mm), x_6 is soil-evaporation (mm)

The following MLR equation was applied on the validation data set of wheat as in Eq. 4:

$$y = -575.0764+0.7262x_1-10.2831x_2+55.0929x_3+ 2.4054x_4-17.4694x_5+5.8133x_6 \qquad (4)$$

The following MLR equation was applied on the testing data set of wheat as in Eq. 5:

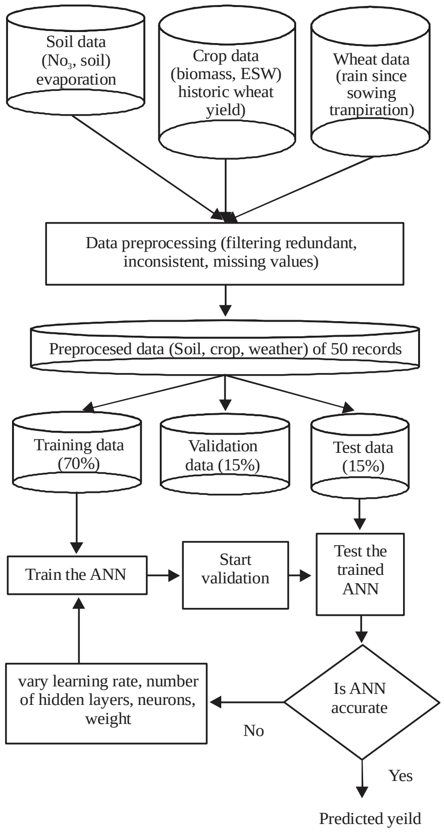

Fig. 3: C-ANN model for wheat yield prediction

Table 4: Steps for implementing MLR

Input: Experimental data set of weather data, crop data and soil data
Output: Predicted wheat yield for the experimental dataset
Step1: Preprocess the data
Step 2: Load the data
Step 3: Fill '1's in the first column of data set and store it in x( )
Step 4: Store yield column in Y( )
Step 5: Use Regress function to apply MLR on the given dataset
Step 6: Display the STATS (1) which gives R square statistics of data

$$y = -4195.8+0.5x_1-5.7x_2+92.4x_3+5.2x_4+17.2x_5-14.7x_6 \quad (5)$$

The regress function of Matlab as used to obtain the model coefficients.

Table 4 shows the steps for implementing MLR on the wheat data set using Matlab R2013a.

RESULTS

Experimental setup: This study explains the experiments carried out on wheat data set using MLR, D-ANN and C-ANN. It also compares the results obtained from them.

The MLR, D-ANN and C-ANN were implemented using Matlab R2013a on windows 7 operating system, Intel® Pentium® CPU P6200 @2.13 GHz processor with 4 GB RAM and 500 GB hard disk.

The accuracy of the three models (MLR, D-ANN, C-ANN) using R^2 statistic and the Percentage Error (PE) was measured. The R^2 is a statistical measure of how close the data are to the fitted regression line[16]. It is given by the Eq. 6:

$$R^2 = 1- (n-1/n-p) (SSE/SST) \quad (6)$$

where, SSE is the sum of squared error, SSR is the sum of squared regression, SST is the sum of squared total, n is the number of observations and p is the number of regression coefficients (including the intercept). The higher the value of R^2, the better is the prediction of the given model. If the value of R^2 is closer to one then it is able to explain most of the variations on the prediction model[16].

The percentage Prediction Error (PE) of a model is computed using the Eq. 7:

$$PE = (|X-Y|/|X|) \times 100 \quad (7)$$

where, X is the actual yield and Y is the predicted yield predicted by the prediction model. The lower the value of PE, the lesser is the error rate and better is the predictive accuracy of the model. Here, R^2 and P.E for the three models MLR, D-ANN and C-ANN models are computed.

Prediction results using default ANN (D-ANN) on wheat crop: The D-ANN consists of one hidden layer. Here, Experiments were conducted by varying the number of neurons in the hidden layer with LR of 0.25 and 0.5.

Table 5 shows the results when D-ANN was applied with 1 hidden layer and a Learning Rate (LR) of 0.25. The neurons in the hidden layer were varied from 20 to 100 and the best result was obtained when the number of neurons in the hidden layer was 60 with a testing R^2 of 0.95.

Table 6 shows the results for 1 hidden layer with LR = 0.5. The neurons in the hidden layer were varied from 20-100 and the best result was obtained when the number of neurons in the hidden layer was 20 with a testing R^2 of 0.91.

From Table 5 and 6, it is observed that the D-ANN model with 20 neurons in the hidden layer with a LR of 0.25 had a higher accuracy of 95% for the test set.

Prediction results using C-ANN on wheat crop: Here, ANN was customized by varying the number of hidden layers to 2.

Table 5: D-ANN results for 1 hidden layer and LR = 0.25

No. of neurons	Training (R^2)	Validation (R^2)	Testing (R^2)
20	0.99	0.25	0.84
30	0.98	0.84	0.74
40	0.99	0.79	0.77
50	0.99	0.75	0.71
60	0.96	0.78	0.95
70	0.99	0.80	0.51
80	0.99	0.45	0.06
90	0.98	0.34	0.73
100	0.1	0.65	0.64

Table 6: D-ANN results for 1 hidden layer and LR = 0.5

No. of neurons	Training (R^2)	Validation (R^2)	Testing (R^2)
20	0.99	0.66	0.91
30	0.82	0.60	0.47
40	0.1	0.81	0.66
50	0.99	0.74	0.62
60	0.88	0.31	0.44
70	0.1	0.86	0.81
80	0.91	0.21	0.60
90	0.1	0.67	0.54
100	0.1	0.59	0.72

Table 7: C-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	0.89
	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	0.42
	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	0.97
	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 8: C-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	0.66	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	0.84
	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	0.85
	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
	90	0.1	0.41	0.69
	100	0.99	0.37	0.43

Within each hidden layer the number of neurons was varied from 20-100. The Customized-ANN (C-ANN) model was tested with learning rates of 0.25 and 0.5.

Table 7 shows the results for 2 hidden layers with LR = 0.25. Keeping the number of neurons in 1st hidden layer fixed (20), the neurons in 2nd hidden layer was varied from 20-100. This experiment was repeated for 20, 30, 40 and 50 neurons in the 1st hidden layer and varied the number of neurons in the 2nd hidden layer from 20-100. This experiment was performed for corresponding neurons in the 1st hidden layer. 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 0.96.

Table 8 shows the results for 2 hidden layers with LR = 0.5. The best result was obtained when the number of neurons in the 1st hidden layer was 20 and the 2nd hidden layer was 30 with a testing R^2 statistic value of 0.93.

Table 9: Prediction results based on R^2 value

Prediction models	Training set (%)	Validation set (%)	Test set (%)
MLR	100	100	92.52
D-ANN	96	78	95
C-ANN	99	90	97

By observing the results (Table 7 and 8), it is observed that the C-ANN model with the following configurations gave the best result of 0.97 (R^2 statistic) for the test set:

- Two hidden layers with 50 neurons in 1st layer and 20 neurons in 2nd layer
- Learning rate = 0.25

Prediction results using MLR on wheat crop: Experiment was conducted using regress function in Matlab. Multiple linear regression was performed on the training set, validation set and the test data set. The R^2 statistic for training set and validation set was 100%. But because of the non-linear relationship between the parameters of the test set the accuracy of the MLR model reduced to 92.52%.

For the training data set of wheat the MLR model coefficients were: $\beta_0 = -1901.8$, $\beta_1 = 0.6$, $\beta_2 = 2.6$, $\beta_3 = -15.6$, $\beta_4 = -0.4$, $\beta_5 = 5.3$, $\beta_6 = 2.3$.

For the validation data set of wheat the MLR model coefficients were: $\beta_0 = -575.0764$, $\beta_1 = 0.7262$, $\beta_2 = -10.2831$, $\beta_3 = 55.0929$, $\beta_4 = 2.4054$, $\beta_5 = -17.4694$, $\beta_6 = 5.8133$.

For the test data set of wheat the MLR model coefficients were: $\beta_0 = -4195.8$, $\beta_1 = 0.5$, $\beta_2 = -5.7$, $\beta_3 = 92.4$, $\beta_4 = 5.2$, $\beta_5 = 17.2$, $\beta_6 = -14.7$

Comparison of D-ANN, C- ANN and MLR prediction models: This section compares the results obtained by the C-ANN model with that of the D-ANN and MLR models on the wheat data set with respect to R^2 statistic and difference between actual and predicted yields.

Accuracy based on R^2 Statistic: Figure 4 shows the comparison of the C-ANN, D-ANN and MLR models on the training set, validation set and test set based on the R^2 statistic.

The accuracies (R^2 statistic values) for the different models are shown in Table 9.

It is observed that the MLR model performed well for the training set and the validation set with an accuracy of 100%. But its accuracy reduced to 92.52% for the test set as it could not capture the non-linear relationship between the different input parameters. The D-ANN model performed with an accuracy of 96% for the training set, 78% for the validation set and 95% for the test set. It is observed that the accuracy for the validation set was less because the D-ANN model failed to generalize.

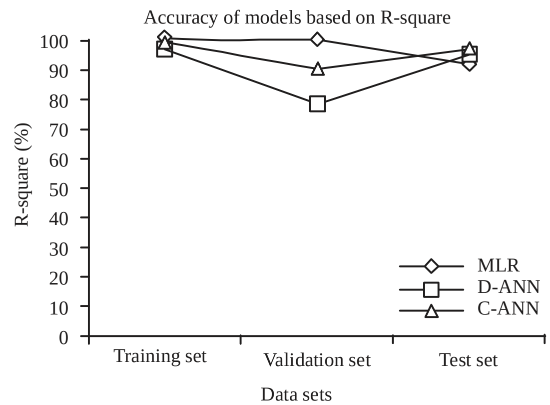

Fig. 4: Comparison of customized ANN and MLR models

Table 10: Percentage error for test set of 8 records for the MLR model

Actual yield (A)	Predicted-yield (Y)	Difference (D = Abs(A-Y))	Percentage error
2224.5	2344.2	119.7	5.381
2359.6	2411.4	51.8	2.1953
2057.3	2166.1	108.8	5.2885
2314.8	2358.9	44.1	1.9051
2284.4	2132.6	151.8	6.6451
3846.8	3754.8	92	2.3916
2735.5	2871.6	136.1	4.9753
3570.6	3399.7	170.9	4.7863

Table 11: Percentage error for test set using D-ANN model

Actual yield (A)	Predicted-yield (Y)	Difference (D = Abs(A-Y))	Percentage error
2224.5	2331.4	106.9	4.8056
2359.6	2392.9	33.3	1.4113
2057.3	2121.3	64	3.1109
2314.8	2335.4	20.6	0.8899
2284.4	2289.6	5.2	0.2276
3846.8	3877.4	30.6	0.7955
2735.5	2787	51.5	1.8827
3570.6	3742.1	171.5	4.8031

The C-ANN model performed consistently well for all the data sets with accuracies of 99, 90 and 97% for the training set, validation set and test sets respectively. The C-ANN model had a higher R^2 statisitic value for the test set compared to the MLR and the D-ANN models.

This indicates that the C-ANN model predicts better for the test set when compared to the MLR and the D-ANN models with an improvement of nearly 2 and 5% over D-ANN and MLR models, respectively.

Accuracy based on percentage error: Table 10 shows the percentage error for test set of 8 records for the MLR model.

Table 11 shows the percentage error for test set of 8 records for the D-ANN model.

Table 12 depicts the difference between actual and predicted yields for test set of 8 records for the C-ANN model.

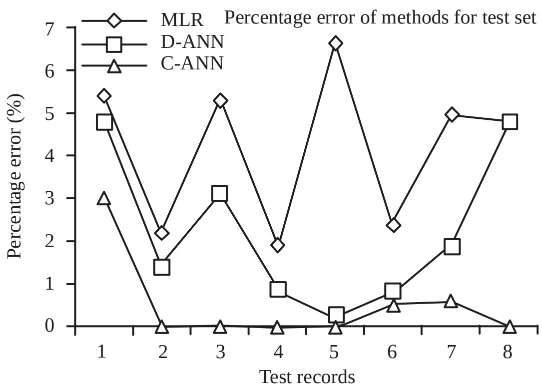

Fig. 5: Percentage error of models for test set

Table 12: Percentage error for test set using C-ANN Model

Actual yield (A)	Predicted-Yield (Y)	Difference (D = Abs(A-Y))	Percentage error
2224.5	2292.8	68.3	3.0704
2359.6	2359.6	0	0
2057.3	2057.3	0	0
2314.8	2314.8	0	0
2284.4	2284.4	0	0
3846.8	3867.4	20.6	0.5355
2735.5	2752.3	16.8	0.6141
3570.6	3570.6	0	0

Table 13: Average percentage prediction errors for each of the models

Parameters	Training set (%)	Validation set (%)	Test set (%)
MLR	4.068	7.053	4.196
D-ANN	2.357	2.5533	2.2408
C-ANN	0.6629	0.3968	0.5275

MLR: Multiple linear regression, D-ANN: Default, artificial neural network, C-ANN: Customized artificial neural network

Figure 5 shows the graph of percentage error for the test set of 8 records for MLR, D-ANN and C-ANN models. Figure 5 depicts that the percentage prediction error for C-ANN model is less than the MLR and D-ANN models. This indicates that the C-ANN model was able to predict better for the test set than the MLR and D-ANN models.

Table 13 shows the average percentage prediction errors for each of the models.

It is observed that the average percentage prediction error for the C-ANN model is lowest for all the data sets. Thus, the C-ANN model is able to predict the yield of wheat better than the MLR and D-ANN models.

DISCUSSION

In this section result of our work is described with results of previous studies conducted in the same area. In Ruß[4] the authors have shown that support vector regression performed with accuracy of 54.92% compared to MLP, Regression Tree and RBF. They considered only RMSE and MAE as performance metrics for accuracy. But in our work we have shown that the proposed C-ANN has a high accuracy of 97%. Also accuracy is measured using different metrics like PPE, R^2. In De Leon and Jalao[6] though the authors have considered many variables for prediction they have considered only JRip algorithm using WEKA tool. They show that JRip performed with a higher accuracy of 89.75% for a full attribute set when compared to reduced attribute set. But our work has proposed a customized ANN model with a greater accuracy. In Ghodsi *et al.*[7] the authors have considered ANN model for wheat prediction. But they have considered many irrelevant parameters like purchasing price which has negatively affected the prediction accuracy. But in this work, critical parameters that directly affect the wheat yield are considered and hence the ANN model was able to perform better. In Ruß *et al.*[9] the authors have mainly considered fertilizer input for wheat yield prediction using ANN. But have not considered rain and other important parameters. Our work may be improved by considering fertilizer also as an input parameter for prediction. In Qaddoum *et al.*[11] the authors have experimented with fuzzy logic for tomato yield prediction. Our work produced better results with accuracy above 90% whereas their work had accuracy in 80's. In Parekh and Suryanarayana[13] the authors obtained good results for training set whereas for test set they obtained very low results. Our work was able to show a better predictive accuracy since the ANN model was customized.

CONCLUSION

Predicting the crop yield is important in agriculture community. In this work wheat yield prediction is done by considering various parameters like Rainfall, transpiration, biomass, Extractable Soil Water (ESW), soil nitrogen (NO_3), soil evaporation and historic wheat yield using ANN and MLR. A comparison study of the results obtained from ANN and MLR is also performed. The results were compared using the R^2 statistic and percentage prediction error. The R^2 values of MLR, D-ANN and C-ANN models on the test set were found to be 92.52, 95 and 97%, respectively. The average percentage prediction error for MLR, D-ANN and C-ANN models on the test set were found to be 4.196, 2.2408 and 0.5275%, respectively. The results indicate that the C-ANN model had a higher R^2 value and a lower percentage prediction error when compared to the D-ANN and MLR models. This shows that the C-ANN model was able to predict the wheat yield better than the MLR and D-ANN models for the given data set.

The outcome of this work may assist the agricultural agencies in providing crop strategies for improving wheat yield.

In future, a generalized prediction model for various crops by considering other parameters like humidity and solar radiation can be developed.

REFERENCES

1. Raorane, A.A. and R.V. Kulkarni, 2012. Data mining: An effective tool for yield estimation in the agricultural sector. Int. J. Emerg. Trends Technol. Comput. Sci., 1: 75-79.
2. McCue, C., 2006. Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis. 1st Edn., Butterworth-Heinemann Ltd., UK., ISBN-13: 978-0750677967, Pages: 368.
3. Dalgliesh, N., M. Saifuzzaman, I. Khan, A.B.S. Hossain and H. Rawson *et al*., 2012. Expanding the area for Rabi-season cropping in southern Bangladesh. Final Report LWR/2005/146, Australian Centre for International Agricultural Research (ACIAR), Australia.
4. Ruß, G., 2009. Data Mining of Agricultural Yield Data: A Comparison of Regression Models. In: Advances in Data Mining. Applications and Theoretical Aspects, Perner, P. (Ed.). Springer, Berlin, ISBN: 978-3-642-03066-6, pp: 24-37.
5. Ramesh, D. and B.V. Vardhan, 2013. Region specific crop yield analysis a data mining approach. Proceedings of the 2nd International Conference on Advances in Computer, Electronics and Electrical Engineering, April 27-28, 2013, Bombay, India, pp: 187-190.
6. De Leon, M.R.C. and E.R.L. Jalao, 2013. A prediction model framework for crop yield prediction. Proceedings of the 14th Asia Pacific Industrial Engineering and Management System (APIEMS) Conference, December 3-6, 2013, Cebu, Philippines, pp: 185.
7. Ghodsi, R., R.M. Yani, R. Jalali and M. Ruzbahman, 2012. Predicting wheat production in Iran using an artificial neural networks approach. Int. J. Acad. Res. Bus. Social Sci., 2: 34-47.
8. Veenadhari, S., B. Mishra and C.D. Singh, 2011. Data mining techniques for predicting crop productivity-A review article. Int. J. Comput. Sci. Technol., 2: 98-100.
9. Ruß, G., R. Kruse, M. Schneider and P. Wagner, 2008. Data Mining with Neural Networks for Wheat Yield Prediction. In: Advances in Data Mining. Medical Applications, E-Commerce, Marketing and Theoretical Aspects, Perner, P. (Ed.). Springer, Heidelberg, ISBN: 978-3-540-70717-2, pp: 47-56.
10. Paswan, R.P. and S.A. Begum, 2013. Regression and neural networks models for prediction of crop production. Int. J. Scient. Eng. Res., 4: 98-108.
11. Qaddoum, K., E.L. Hines and D.D. Iliescu, 2013. Yield prediction for tomato greenhouse using EFuNN. ISRN Artif. Intell., Vol. 2013. 10.1155/2013/430986
12. Singh, R.K. and Prajneshu, 2008. Artificial neural network methodology for modelling and forecasting maize crop yield. Agric. Econ. Res. Rev., 21: 5-10.
13. Parekh, F.P. and T.M.V. Suryanarayana, 2012. Impact of climatological parameters on yield of wheat using neural network fitting. Int. J. Modern Eng. Res., 2: 3534-3537.
14. Jain, A.K., J. Mao and K.M. Mohiuddin, 1996. Artificial neural networks: A tutorial. IEEE. Comput., 29: 31-44.
15. Shedden, K., 2004. Multiple linear regression. Department of Statistics, University of Michigan, USA. http://dept.stat.lsa.umich.edu/~kshedden/Courses/Stat401/Notes/401-multreg.pdf
16. Zhang, N.R., 2010. Lecture 5: Multiple linear regression. Statistics 203, Stanford University. http://statweb.stanford.edu/~nzhang/203_web/lecture4.pdf