Effectiveness of Proposed New Machine Learning Algorithm ReANN with ANN using Agriculture Data

D.Maghesh Kumar
Research Scholar
PG & Research Dept. of Computer Science
Rajah Serfoji Government College, Thanjavur
Affiliated to Bharathidasan University, Trichy
Tamil Nadu, India

Dr. K. Mohan Kumar
Research Supervisor & Head
PG & Research Dept. of Computer Science
Rajah Serfoji Government College, Thanjavur
Affiliated to Bharathidasan University, Trichy
Tamil Nadu, India
For Correspondence: tnjmohankumar@gmail.com

***Abstract*- The economy of India chiefly depends on agriculture and agro related industrial products. In agriculture, the technological development is inevitable and India is rapidly moving towards it. The farmers are beneficial, when using this technology to increase crop productivity. In this paper, the machine learning algorithm artificial neural network (ANN) and the proposed reengineered artificial neural network(ReANN) has been analyzed and compared. The findings of this work, proves that the reengineered artificial neural network model provides better result than artificial neural network model.**

***Keywords*- agriculture, analyze, artificial neural network, crop yield, machine learning.**

I. INTRODUCTION

Agriculture has always been one of the vital occupations that serve mankind, both in terms of livelihood and employment. Due to the substantial increase in the population, there has been an immense gap between the demand and supply of food and edible resources. All the countries in the world try to achieve extraordinary increase in the crop production with the help of expansion of irrigation and technological innovation in agriculture. There are multiple strategies which can be adopted to improve the number and quality of crops [1].

The agricultural yield is primarily depends on weather conditions, which also influences the crop yield but weather conditions are beyond human control. From the beginning of the history of agriculture, crop yield forecasting has been a biggest worry for the farmers. Forecasting techniques have evolved, as has agriculture itself and the specifications of the forecasts needed [2]. In this context, the farmers necessarily require a timely advice to predict the future crop productivity and to maximize the crop production in their crops.

Scientists all over the world have done much research on developing different types of yield estimating models. However, due to the complexity and nonlinear character of agricultural ecosystems, problems existed in the model using traditional statistical algorithms [3]. In the recent studies, researchers have recognized that nonlinear models are more realistic and potentially more accurate. However, many of them are overly simplistic because it is difficult to know the appropriate nonlinear forms of cropland ecosystem functions [4][5]

A. *Artificial Neural Network (ANN)*

ANN is a prediction technique that is used to predict non- linear relationships from an input provided[6]. The most popular ANN architecture is a three-layered, which comprises the input, hidden and output layers[7]. An experiment is run by giving the input values as attributes. The weights and activation function values are added and applied at each node of the hidden layer, till the desired output of the feed forward algorithm is achieved. The output at every level is compared with the expected output from the data set. If the output doesn't match, it is sent back to the initial layer and is given different values for weights and activation function and then back propagation algorithm is applied [8]. The method finally stops when it matches with the expected output.

During training, the network learns by adjusting the weights connecting the input and hidden layer and the weights connecting the hidden layer and output, by the learning rate parameter. The comparison of the current predicted value with the known value is calculated by root mean square error (RMSE) which was then propagated back through the feed-forward links[9]. The back-propagation algorithm modifies the weights of all internal links to minimize the error. Finally, after all epochs of the training phase were processed the network was ready for prediction.

The back-propagation learning algorithm, is characterized by two parameters, the training rate and the momentum were used by the process in progressive refinement of weights. The major advantage of ANN is that the training, which is the most time consuming exercise, is done off-line[10].

II RELATED WORK

K. Menaka and N.Yuvaraj discussed various successful methods used for predicting and modeling crop yields. It is very difficult to find dependable attribute and these attributes will play a crucial role in building crop yield model[11].

G. Ruß and R. Kruse say several crop yield prediction datasets have convincingly with a small number of attributes and an meticulous approach can be applied for the purpose of comparing the model [12].

G.Sanchez, J.Solis, and W. Ojeda states impact of attribute selection in prediction models and the necessity of separating the dataset into training and validation datasets. In the development of prediction models, major part of the dataset was used for training and the rest of the dataset was utilized for validation [13].

Shastry, K. A., Sanjay, H. A., & Deshmukh, A customizes an ANN by varying the amount of hidden layers(HL), Learning Rate (LR) and a number of neurons in the hidden layer and by proposing a new model for crop yield forecasting. In their work they compare the proposed model with ANN and Multiple Linear Regression (MLR) models by means of the prediction accuracy in wheat yield [14].

R. Beulah studied various data mining techniques such as Bayesian Network, ANN, Partial Least Squares Regression, and Support Vector Machine for predicting crop yield. The merits, demerits and performance metrics of these techniques developed by previous researchers are tabulated [15].

B.Ji, Y. Sun, S.Yang And J.Wan developed an easy and precise assessment technique to envisage rice yields in agricultural management planning process and the efficiency of multiple linear regression models with ANN models are calculated. By the comparing two models, ANN model is more accurate method for crop yield prediction [16].

Snehal S.Dahikar, Dr.Sandeep V.Rode proposed ANN to reduce losses when the conditions are not favorable while predicting the crop yield and based on the parameters of the soil, the various atmospheric conditions, guaranteed price, area under cultivation, subsidy, insured area etc.. The proposed ANN model is more efficient in estimating long - term or short-term crop production[17].

M.Odisa et. al. presented, the crops used were cotton, groundnut, rice, soya bean, corn, wheat etc. and the parameters such as soil type, PH, nitrogen, depth, temperature and rainfall were taken into account for estimating the maize yield prediction by using ANN models in the major maize producing provinces of South Africa is presented [18].

Nafiseh, Yaghmaeian, Mahabadi explored in detail about the working of ANN models with change in the count of neurons in hidden layer and back propagate the model for optimization. Varying the parameters such as learning rate, epoch increases the capability of the model to foresee crop yield efficiently. In the ANN models, number of hidden nodes must be kept at optimum level otherwise it results in overfitting and Root mean square error(RMSE), was used to measure the accuracy of created model[19].

III MATERIALS AND METHODS

A. *Data source and description*

For this research, last 20-year data is collected from the concerned departments of Statistics and Agriculture, National Informatics Centre, imdchennai.gov.in, tnagrisnet.tn.gov.in, data.gov.in and merged into one, such that a dataset sufficiently large enough for the study is created.

Initially split the dataset into two parts namely Train and Test and further split into a training set and a validation set. Validation set is the subset of the training dataset, then the model is iteratively trained and validated on these different sets is commonly known as Cross Validation. Test set used only to assess the performance of a fully-specified model.

B. *Input variable selection*

In general, one of the difficulties faced in the prediction process is that most of the essential parameters that are necessary to consider for the accurate prediction are not considered. This reduces the efficiency of the predicted results which in turn leads to lack of proper forecasting.

Prediction can be done either linear way or nonlinear way; linear models are inefficient for prediction. Whichever model is chosen, each independent variable is first checked for its significance with the dependent variable. So, filter out the variables from the whole data set based on expert advice, then eliminate variables and attributes which are in non-numeric format. Next remove the variables which have identical properties/values but measured in different format/units and choose to experiment with more standard measuring unit.

Three important factors were considered for crop yield models are crop type, Soil Properties, and environmental factors, such as climatic information [20]. Selected parameters for the crop yield are depicted in the Table I.

TABLE I: PARAMETERS FOR CROP YIELD

Dataset	Parameter	Description
Weather	Rainfall(mm)	Amount of rainfall
	Avg.Temp(C)	Average Temperature
Soil	Soil Water Content (mm)	Amount of water in a soil
	Evapotranspiration(mm)	Sum of evaporation from the plant and transpiration from the Soil.
Crop	Cultivated Area hectare)	Area under cultivation
	Production (Tones/ hectare)	Crop production for the cultivated area
	Crop Yield (Kg/ Hectare)	Obtained by dividing the production by the area cultivated.

The main idea behind this research is to deliver best envisaging and analyzing techniques to attain best possible results and the most common data-driven techniques applied to Crop Yield Prediction were selected for this work is ANN[7][8][14][18].

C. *ANN Algorithm*

The ANN algorithm is described as follows

Input: *Selected input variables from the crop, weather and soil data from the initial raw dataset is taken as input.*

Processing Steps:

Step 1: *Randomly initialize the weights of ANN model*
Step 2: *Train the ANN model on the training set*
Step 3: *Evaluate each trained model's performance based on the validation set*
Step 4: *Evaluate the trained, cross validated ANN model on test set*
Step 5: *Obtain and evaluate trained models prediction accuracy and prediction error*
Step 6: *Find the deltas for the preceding layers by propagating error backwards to adjust the weights.*

Output: *Predicted yield.*

D. Reengineered Artificial Neural Network

In this work, the ReANN model is done by hyper-parameter optimization i.e. Optimize the values of Learning Rate, Number of Epochs, Hidden Layers, Hidden Units, momentum and Activations Functions. The methodology followed for reengineering the ANN is depicted in following ReANN algorithm used to predict crop yield.

ReANN Algorithm

Input: *Selected input variables from the crop, weather and soil data from the initial raw dataset is taken as input.*

Processing steps:

Step 1: *Taking care of missing, inconsistent and remaining redundant data from the initial dataset for further processing.*
Step 2: *Split the preprocessed dataset into train, validation and test sets in 60-20-20 ratio.*
Step 3: *Train the ANN model using damped-least squares method.*
Step 4: *Back propagate the ANN using Leaky ReLU activation function for hidden layers and for output layer using softmax function*
Step 5: *Cross validate the trained model's performance on the validation set.*
Step 6: *Check the trained, cross validated ANN model with test set.*
Step 7: *Evaluate trained models prediction accuracy and prediction error.*
Step 8: *Fine tune the trained ANN model using hyper-parameter optimization.*
Step 9: *Go to step 3 and repeat the steps until obtaining the maximum prediction accuracy with minimum prediction error.*

Output: *Predicted yield.*

The model of the methodology has been depicted in the fig.1. Initially the crop, soil, weather details were collected and needed parameters were selected and grouped together as a single dataset. The initial dataset is cleaned by removing missing values and outliers. The processed dataset has been split into 60-20-20 ratio.

The procedure used to carry out the learning process in a ANN is called the optimization. The objective of learning problem is to the minimize loss index, which measures the performance of a neural network on a data set and used to prevent overfitting, by controlling the effective complexity of the ANN.

The damped least-squares method, also known as the Levenberg-Marquardt algorithm, has been designed to work specifically with loss functions. This method makes it to be very fast when training neural networks measured on that kind of errors.

Leaky ReLU is used to do back propagation in the hidden layers of ANN. In Leaky ReLU where slope is changed left of x=0 and thus causing a leak and extending the range of ReLU. Instead of the function being 0 when x < 0, a leaky ReLU will instead have a small negative slope (of 0.01, or so).

Softmax is a very interesting activation function used in the output layer back propagation because it not only maps our output to a [0,1] range but also maps each output in such a way that the total sum is 1. The output of Softmax is therefore a probability distribution.

Hyper parameters are tuned by running the whole training job, looking at the aggregate accuracy, and adjusting to find the best combination to handle the problem. The hyper parameters used in the ANN are learning rate, momentum and epoch. The process shown in fig.1 will be repeated until getting accurate yield prediction. The ANN and ReANN models have been implemented using Python.

E. Performance Measure

Performance of the models are measured using two metrics namely RMSE and Percentage Error(PE).

RMSE helps to expect how our model to be on its next prediction by evaluating trained model`s accuracy. In general, a lower RMSE is better than a higher one. To calculate the RMSE, the following equation is used

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - o_i)^2} \qquad \text{-----} \quad (1)$$

Where, n is number of samples, f is forecasts and o is observed values.

PE is calculated using the following formula

PE = ----- (2)

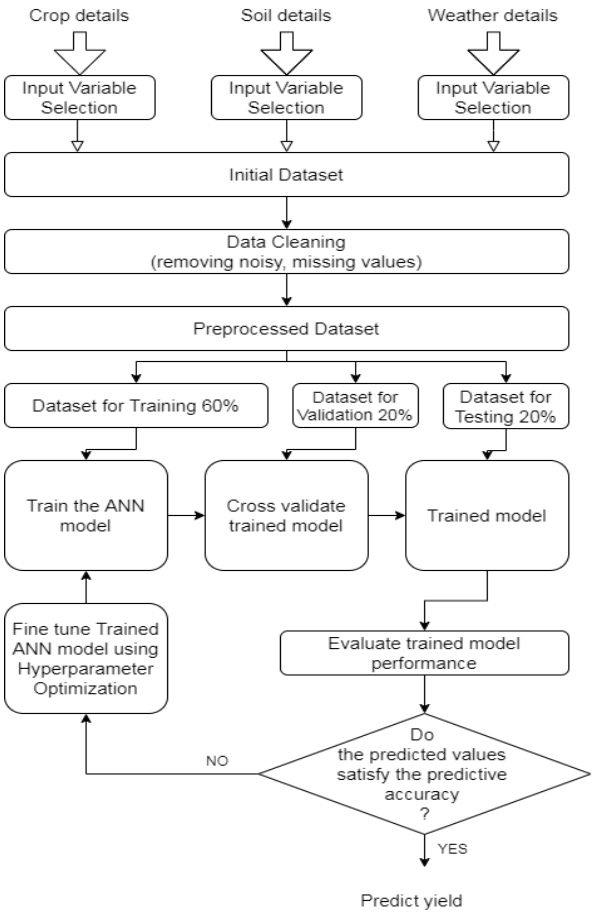

Fig 1: ReANN Model for yield prediction

IV RESULTS AND DISCUSSION

The ultimate objective of this work is to help the farmers by forecasting the yield of crop in Tamil Nadu Dataset used in this research consists of 7 input variables depicted in table 1 and crop yield which is a target variable. To analyze the dataset, ANN and ReANN algorithms were used for analysis. The training set is 80% which comprises the data from the year 1999-2014, the validation set 20% is taken from the training set and testing set is 20% which is from the year 2015-2018 data.

Usually a threshold value of the error or a fixed number of epochs (minimum of 1000 epochs) is considered as the stopping criteria.

A. Performance evaluation

While evaluating the ANN model, learning rate (LR) is 30 and no. of hidden layer is one. The Forecasted yield(FY) is predicted using ANN from actual yield(AY) and the PE is calculated using the formula in equation 2. The forecasted yield and PE is shown in the Table II. The comparison of forecast yield with actual yield is visualized in fig. 2.

TABLE II : YIELD FORECASTING USING ANN AND REANN

Year	AY	FY (ANN)	FY (ReANN)	PE (ANN)	PE (ReANN)
1999	663	663	665	0.00	0.30
2000	612	593	601	3.10	1.80
2001	856	839	845.66	1.99	1.21
2002	908	891	901.56	1.87	0.71
2003	1301	1244.45	1284.33	4.35	1.28
2004	1428	1450.77	1465.23	1.59	2.61
2005	998	982	990	1.60	0.80
2006	1190	1173	1185.67	1.43	0.36
2007	1530	1457.26	1484.69	4.75	2.96
2008	1564	1547	1569	1.09	0.32
2009	1016	1018	1016.89	0.20	0.09
2010	1343	1303	1325.77	2.98	1.28
2011	1281	1235.03	1254.56	3.59	2.06
2012	916	889.8	899.21	2.86	1.83
2013	1037	1040.44	1043.59	0.33	0.64
2014	935	910.22	919.34	2.65	1.67
2015	1122	1053.29	1072.91	6.12	4.38
2016	1087	1062.54	1074.53	2.25	1.15
2017	1377	1343	1358.23	2.47	1.36
2018	1484	1435.03	1462.31	3.30	1.46

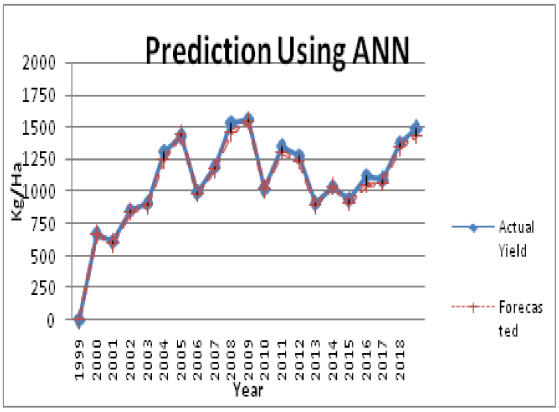

Fig 2: Actual yield vs. Forecasted yield by ANN

The ReANN model is setup its hyper parameters as given in table III to fine tune the model.

TABLE III: PARAMETERS USED FOR REANN

Parameters	**Chosen Value for ReANN**
Learning Rate	0.5
Momentum	0.3
Hidden layers	2
Training Time	500
Epochs(Fixed)	1000

The PE is calculated using the formula in equation 2. The FY of the crop is predicted using the ReANN model and PE are shown in the Table 2. By observing the fig. 3, the forecast yield is nearly close to actual yield.

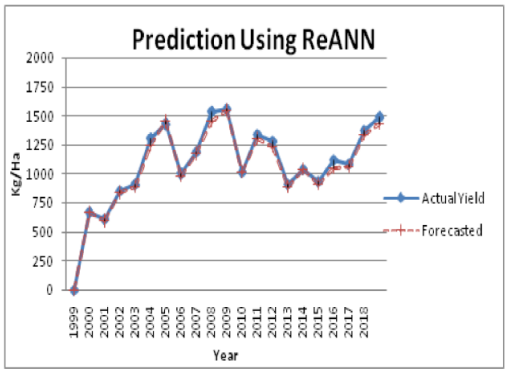

Fig 3: Actual yield vs Forecasted yield by ReANN

Prediction of the two models is compared by its PE is shown in fig.4**.** By interpreting the diagram, it states ReANN has less prediction error than ANN.

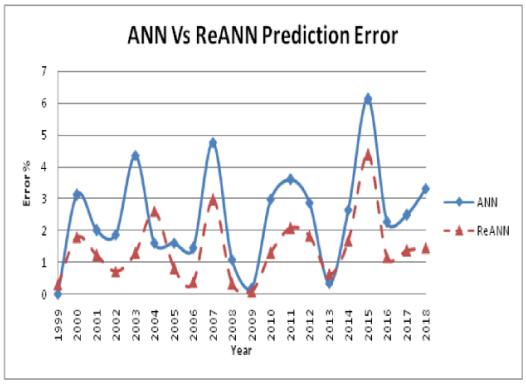

Fig 4: ANN vs. ReANN prediction error

The RMSE values are computed using the formula in equation 1. From the fig. 5 the reader can see the deviation of observed value from the model value of two models.

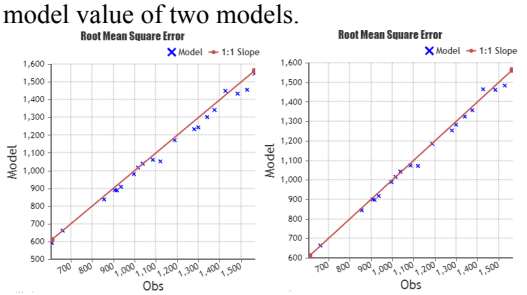

Fig 5: RMSE of ANN and ReANN

The RMSE values of training, validation and testing phases of the two models are tabulated in table IV and the comparison of the RMSE values are shown in fig. 6. The ReANN model has less RMSE error value than ANN in all the three stages means, this measure of the metric shows ReANN model has more accuracy then ANN.

TABLE IV. COMPARISON OF THE RESULTS FOR THE RMSE

Model	Training(%)	Validation (%)	Testing (%)
ANN	32.125	29.264	35.131
ReANN	19.139	17.816	21.284

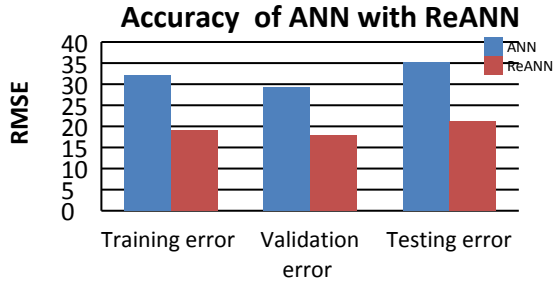

Fig 6: Comparison of ANN and ReANN models

V CONCLUSION AND FURTHER ENHANCEMENT

This research is to help farmers and agro industry for making better decision in crop yield. Crop data is collected from the web resource for exploratory data analysis and prediction was done with ANN and ReANN models. After extensive tuning of hyper parameters for the ReANN model, the prediction results of the models ANN was compared with ReANN model. The accuracy results of the ReANN model shows, improved performance in yield prediction over ANN.

The onset of monsoon affected germination and initial growth, whereas, abnormal weather conditions affected quality of produce. The combined effect of rainfall and temperature was on the relative humidity that created a conductive atmosphere for insect and pest attacks on crops. In addition to the weather and soil parameters, it would be natural to include fertilizer and pesticides information as exogenous inputs. This will envisage more accurate crop yield and be a topic for future research.

REFERENCES

[1] Wu Fan, Chen Chong, Guo Xiaoling, Yu Hua, Wang Juyun, "Prediction of crop yield using big data", *8th International Symposium on Computational Intelligence and Design (ISCID)*.2015; 1, 255-260.

[2] Teresa Priyanka, Pratishtha Soni, C. Malathy, "Agricultural Crop Yield Prediction Using Artificial Intelligence and Satellite Imagery", *Eurasian Journal of Analytical Chemistry*, 2018 13 (SP): 6-12.

[3] Kimes, D. S., Nelson, R. F., Manry, M. T., And Fung, A. K., "Attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements".. *International Journal of Remote Sensing,* 1998, 19, 2639–2663.

[4] Rasmussen, M. S., "Developing simple, operational, consistent NDVI-vegetation models by applying environment and climatic information. Part II: Crop yield assessment", *International Journal of Remote Sensing*, 1998,19, 119–139.

[5] Hayes, M. J., and Decher, W. L., "Using NOAA AVHRR data to estimate maize production in the United States Corn Belt", *International Journal of Remote Sensing*, 1996, 17, 3189–3200.

[6] Elsner, J. B., and A. A. Tsonis. "Nonlinear prediction, chaos, and noise." *Bulletin of the American Meteorological Society* 1992,73, no. 1, 49-60.

[7] Jeeva C, Shoba S.A.,"An Efficient Modelling Agricultural Production Using Artificial Neural Network (ANN)", *International Research Journal of Engineering and Technology (IRJET)*, May-2016, Volume: 03 Issue: 05.

[8] Meena, Mukesh, and Pramod Kumar Singh. "Crop Yield Forecasting Using Neural Networks." *In International Conference on Swarm, Evolutionary, and Memetic Computing*, *Springer, Cham*, 2013, pp. 319-331.

[9] Jiang, Yingni., "Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models"., *Energy policy 36, no. 10,*2008, 3833-3837.

[10] Szkuta, B. R., L. Augusto Sanabria, and Tharam S. Dillon. "Electricity price short-term forecasting using artificial neural networks." *IEEE transactions on power systems,* 1999, 14, no. 3, 851-857.

[11] K. Menaka N.Yuvaraj, "A survey on crop yield prediction models", *Indian Journal of Innovations and Developments,* October, 2016, Vol 5(12),

[12] G. Ruß and R. Kruse, "Feature selection for wheat yield prediction,", *Research and Development in Intelligent Systems XXVI, M. Bramer, R. Ellis, and M. Petridis, Eds., Springer, London, UK*, 2010, pp. 465–478.

[13] Gonzalez-Sanchez, A., J. Frausto-Solis, and W. Ojeda-Bustamante. "Attribute selection impact on linear and nonlinear regression models for crop yield prediction", *Sci. World J*. 2014.
[14] Shastry, K. A., Sanjay, H. A., & Deshmukh, A. "A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction", *Journal of Artificial Intelligence*, 2016,vol,9, pp. 23-32,.
[15] R. Beulah, "A Survey on Different Data Mining Techniques for Crop Yield Prediction", *International Journal of Computer Sciences and Engineering*, Jan 2019, Vol.-7, Issue-1.
[16] B.JI, Y. SUN, S.YANG and J.WAN "Artificial neural networks for rice yield prediction in mountainous regions". *Journal of Agricultural Science* ,2007, 145, 249–261.
[17] Snehal S.Dahikar, Dr.Sandeep V.Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", *International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering*, 2014,2(1), 683-686.
[18] Omolola M. Adisa ,Joel O. Botai, Abiodun M. Adeola , Abubeker Hassen et. al "Application of Artificial Neural Network for Predicting Maize Production in South Africa", *Sustainability*, 2019, 11(4), 1145.
[19] Nafiseh, Yaghmaeian, Mahabadi, "Use of the Intelligent Models to Predict the Rice Potential Production", *International Academic Journal of Innovative Research*, 2015,Vol. 2, No. 10, pp. 14-25.
[20] Naheed Ejaz, Shabbir Abbasi, "Wheat Yield Prediction Using Neural Network and Integrated SVM-NN with Regression", *Pakistan Journal of Engineering Technology and Science (PJETS),* December 2018,Volume 8, No 2.