

Studying the Impact of Water Supply on Wheat Yield by using Principle Lasso Radial Machine Learning Model

Muhammad Adnan

Institute of Manufacturing Information and Systems,
Department of Computer Science and Information
Engineering, National Cheng Kung University,
Tainan, Taiwan

Naheed Akhter

Department of Computer Science,
GC University Faisalabad,
Pakistan

M. Abid, M. Ahsan Latif, Abaid-ur-Rehman

Department of Computer Science,
University of Agriculture Faisalabad,
Pakistan

Muhammad Kashif

Department of Computer Science,
Allama Iqbal Open University,
Islamabad, Pakistan

Abstract—Wheat plays a vital role in the food production as it fulfills 60% requirements of calories and proteins to the 35% of the world population. Owing to wheat importance in food, wheat demand is increasing continuously. Wheat yield is committed to the availability of water supply. Due to climatic and environmental variations of different countries, water supply is not available in constant and desire quantity that is necessary for better wheat yield. So, there is a strong relationship and dependency that exists between water supply and wheat yield. Therefore, water supply is becoming an issue because it directly effects wheat yield. In this research, a Principle Lasso Radial (PLR) model is proposed using Machine Learning technique to measure the effect of water supply on wheat yield. In this Principle Lasso Radial (PLR) model, various experiments are conducted with respect to the performance metrics, i.e. relative water contents, waxiness, grain per spike and plant height. Principle Lasso Radial (PLR) model's produced reduced dimensional data with respect to performance metrics. That data is provided to Radial Basis Neural Network (RBNN), and it showed regression values R under different water supply conditions. Principle Lasso Radial (PLR) model achieved an accuracy of 89% among variance Machine Learning techniques.

Keywords— Radial basis function (RBF); Radial Basis Neural Network (RBNN); ANN; lasso; principle component analysis (PCA)

I. INTRODUCTION

It will predict wheat yield under stress and normal conditions by using ML. This model will enable us to predict the factor measurements which are necessary to have the required yield in future. The dependency of attributes of the wheat will be measured by using ML. The proposed model will predict the growth under stress and normal condition. Genotype with 1600 sets under different conditions will be tested by ML model. These sets of 1600 different values will be presented varyingly under normal and stress conditions.

The ML model will predict and tell the yield which can be obtained by using different features under these conditions. The model will map and find regression and classify these attributes with respect to wheat yield.

The Machine learning technique enables us to find both the anticipatory as well as relational problems. This technique is applied to analyze the water effect on wheat crop and traits for better agriculture production. The machine learning algorithm is applied for the classification of yield component in order to get high yield. Wheat is the key and basic part of the food. It fulfills and provides a large number of calories and vitamins to human beings. It is a large element of food. Because of its significance, its demand has been enlarged for the last many decades. It is very fundamental to fabricate high yield of wheat to meet the supplies. Wheat also contains critical amino acids, minerals, and vitamins, beneficial photochemical and dietary fiber apparatus essential for the human diet, and these are predominantly enriched in whole-grain foodstuffs. It is not only a food factor but it also contains many proteins, fats and power in it which are vital and positive for human body [1].

The Agriculture sector in Pakistan is a major source of livelihood; 21% of Growth Domestic Production depends on it. Country's 44% workforce is related to this sector. Wheat is the essential part of food in Pakistan. Therefore, it grows almost in all parts of the country. As the population is increasing rapidly, there is need of growing wheat at an extensive level. The population is increasing with the rate of 3.1% yearly actions [2]. The area specified for wheat production is more than any other crop. This is because its consumption is more than any other crop. It requires a specific amount of water and minerals for growth [3].

Many environmental factors like the presence of Salt in the soil, water stress caused by lesser water in the plant and waterlogging poorly affect plant development. The climatic

change also contacts the plant growth and has an important impact. It is not possible for a plant to set itself fully free from ecological circumstances because of environmental variation habit [4]. Plant's cells, genetics, biophysics, nutrition, photosynthesis, fertility rate and hormones are affected by it. Typically, reaction and environmental acceptance rate of different plants are different from one another. Not all the plants act the same way under different growth situations. Some plants can face unfavorable environment whereas the others cannot do this [5].

II. RELATED WORK

The phenomenon of stress alteration can be seen in two ways that are stress tolerance and stress avoidance. Both have different perspective and strategy about stress. In avoidance mechanism, plants grow speedily and become more talented to pass up stress collision. as the plant is at the peak level of its life, so it can easily accomplish at its final growth stage and can avoid stress environment. On the other hand, the patience method is little risky and dominant. At tolerance level, plants are taught by Nature to live under diverse environmental and stress full conditions. Even though in this plan, plants have to negotiate on their growth and enlargement procedure [6].

Dahikar and Rode [7] analyzed the crop prediction on the basis of different environmental conditions. Demographical conditions, temperature, and many other factors affected the prediction system. It was important to understand the environmental effect because crop production was directly affected by this phenomenon. They used feed-forward back-propagation ANN model to discover the crop yield under different stress conditions including water, ionic, weather and many other parameters like type of soil, PH, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulphur, manganese, copper, iron, depth, temperature, rainfall, humidity. This network can take a choice on the source of the training set of data which has been provided to it.

Emamgholizadeh et al. [8] premeditated that in the agricultural research the most vital purpose of breeding was the production of seed yield. In this research, two techniques were used; ANN and Multiple Regression Models (MLR). Both techniques were used to forecast same seed yield on the basis of premeasured features of the plant like maximum flowering days, the height of the plant in centimeters, numbers of capsules of each plant, etc. The results were experienced by using both MLR and ANN techniques and findings showed that ANN was more correct with respect to the root mean square error and founded coefficient.

Adnan et al. [12] use the machine learning methods to measure the evapotranspiration rate. Evapotranspiration diverges with the climate change and as the climate has a lot of deviation geographically, the pre-developed systems have not used all available meteorological data hence not robust models. In their study, a model is developed to evaluation evapotranspiration with extra authentic and accurate reduced meteorological parameters using different machine learning methods. The dataset with reduced dimension is modeled through time series neural network. Time series neural network delivered better results as compared to other existing approaches.

III. CONTRIBUTIONS

In ML, there are many approaches that can be used for the best analysis of the water stress. But the current study used Stepwise regression and Principle component analysis (PCA). Both the techniques are suitable for reducing the dimension. By reducing the data set, the well dependent elements can be obtained which will make the analysis easy. Then the reduced dependent data will be analyzed by using the Radial Basis Neural Networks (RBNN). Some work related to this use of machine learning approach has been performed. Water factor was studied deliberately and very critically.

IV. EXPERIMENTAL SETUP

Experimental material: The data related to wheat crop was collected in two years. In this procedure, the plant was grown in four different seasons. This was done at the department of Plant Breeding and Genetics, University of Agriculture, Faisalabad. The area where the selected population was grown consisted of a 5-meter wide long plot and with 15 cm and 30 cm plant to plant and row to row distances, respectively. The plant production procedure was examined under stressed and normal conditions. Under normal condition, the water quantity provided to the land was sufficient as according to the area allocated. When the growth of wheat plant was analyzed under the stressed condition, the quantity of water was declined and not enough water necessary for growth was supplied. By reducing the water supply or creating an artificial shortage of water, plant growth was analyzed.

The data for heading, relative water content (RWC) and proline were recorded before anthesis while peduncle length (PL), extrusion length (EL), awn length (AL), plant height (PH), seed size (SZ), grains per spike (gr/s), and yield/plant (Y) were recorded at maturity. The machine learning techniques were applied to the collected data. It was meant to find out the relation between the yield and the water-stressed conditions. Following three approaches were used for this purpose.

A. Principle Component Analysis (PCA)

Principle Component Analysis can simplify the problem by replacing a group of variables with a single new variable. The principal component analysis is a quantitatively rigorous method for achieving this simplification. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information. The first principal component is a single axis in space. The first plotted line on the graph represents the greater variation among all the data sets. When projected each observation on that axis, the resultant values form a new variable. And the variance of this variable is the maximum one among all possible choices of the first axis.

$$P_e = \sum_{k=1}^n dkE_{ke} \quad (1)$$

This is how our data will process using this equation, where "e" is the principal component. All our traits will be treated as separate principal components. Here in (1) "Pe" will be output generated for each component. "k" is the input data

set that is related to the specific component. In this case, each trait will be treated as a separate input band. “n” is the total number of input values. Here the PCA, during its process, calculates the eigenvector value that is represented by “E”.

The second principal component is another axis in space, perpendicular to the first. Projecting the observations on this axis generates another new variable. The variance of this variable is the maximum among all possible choices of this second axis. The full set of principal components is as large as the original set of variables. But it is commonplace for the sum of the variances of the first few principal components to exceed 80% of the total variance of the original data [9].

Step 1 - Standardize: The data that is collected from the field is not able to process our function directly. It needs ‘new’ in a ready form so that we can obtain a precise and efficient result. The data units are not same for every data component during the collection process. So before processing, the data must be standardized into an acceptable form, in our case that is performed in PCA Matlab function.

Step 2 - Calculate covariance: Covariance is a measure of how much random variables change together. In our data, the variables are large in quantity. The value of covariance among all those will tell about the actual relation of data with the outcome.

$$\sum n_i = 1(X_i - \bar{x})(Y_i - \bar{y}) n-1 \quad (2)$$

With \bar{x} and \bar{y} denoting the means of X and Y, respectively. X denotes the input variable and Y denotes the output variable. All these above following variable are treated as Y. The covariance among the variables is shown below.

Step 3 - Selecting Principal Components: PCA reduced the data dimension on the basis of dependency or variation of variables. In our data, PCA will reduce the eleven traits into a specified number of trait variables on the basis of their dependency on yield that is our outcome.

$$\Sigma v = \lambda v \quad (3)$$

Where in (3),

Σ = Covariance matrix , v =Eigenvector, λ =Eigen value.

Here each trait will be principle component and will be decided by the eigenvectors directions on axis since they have all the same unit length. We can check that the eigenvectors, the Eigen-value calculation is correct through the equation.

Step 4 - Transforming the samples onto the new subspace: In the last step, we use-dimensional matrix W that we just computed to transform our samples onto the new subspace via the equation. These new traits were used for future estimation of the wheat yield.

$$Y = Wt \times X \quad (4)$$

B. Lasso and Elastic Net

Lasso is a regularization technique for performing linear regression. Regularization is a technique used excessively to solve the compound problems. Lasso has the capability and

functionality to reduce the dimensions of the data. Therefore, it resembles ridge regression. Lasso in its property can compress the variables into a suitable number of values and this property makes it a shrinkage estimator. It produces small coefficient which acts as an estimator in the data processing. In lasso method, the estimator produces lesser and minimum value of mean square error. While on another hand, when a generalized least squares estimator is applied, it produces a large value as compared to lasso estimator.

As a number of penalty terms increases, lasso coefficient values become closer to zero. This technique is different from ridge regression. This means that the lasso estimator is a smaller model, with fewer predictors. As such, the lasso is an alternative to stepwise regression and other model selections and dimensionality reduction techniques [10].

Elastic net is a lasso related technique. It is a hybrid of ridge regression and lasso regularization. It is a kind of mixture of both. Like lasso, the elastic net can generate reduced models by generating zero-valued coefficients. Empirical studies have suggested that the elastic net technique can outperform lasso on data with highly correlated predictors.

The lasso technique solves this regularization problem. For a given value of λ , a non-negative parameter, lasso solves the problem

$$\min_{\beta_0} \beta \left(\frac{1}{2N} \sum_{i=1}^N \left(y_i - \beta_0 - x_i^T \beta \right)^2 + \lambda \sum_{j=1}^p |\beta_j| \right) \quad (5)$$

Where in (5) “N” is the number of observations. Observation is the quantity of our data which has been collected during the field experiment. “ y_i ” is the response at observation. It means that against every observation, we got some related response when it was made. “ x_i ” is data. X_i is data values related to our traits. It will be represented as a vector of p values at observation i. “ λ ” is a positive regularization parameter corresponding to one value of Lambda. The parameters β_0 and β are scalar and p-vector, respectively.

The elastic net technique solves this regularization problem. For a α strictly between 0 and 1, and a nonnegative λ , elastic net solves the problem

$$\min_{\beta_0} \beta \left(\frac{1}{2N} \sum_{i=1}^N \left(y_i - \beta_0 - x_i^T \beta \right)^2 + \lambda P_{\alpha}(\beta) \right) \quad (6)$$

Here, the parameter is same as the lasso when the value of α remains 1. “ α ” is the processed outcome of the entire data that is given during the formalization.

C. Data Modeling

Another approach used in our work was that of Data Modeling. It was applied to a data set that was collected after obtaining the result of the PCA and the stepwise regression. Radial basis neural network is the next type of the machine learning method that is very useful in finding the regression among different variables. Radial basis neural network is based on the number of layers and activation, training function

with a specific number of neurons in training of a data. It is used because of its rapid training and learning process. It provides good interpolation of the data set as compared to other neural networks [11].

$$a = \text{radbas}(\|W \cdot p - b\|) \quad (7)$$

Here in (7) the network output is the output layer consisted of a single neuron. The activation function used here is “radial bias” and it is made of weight W and input vector product with b bias.

V. RESULTS AND DISCUSSION

We provided data set in the principal component analysis method showing in Fig. 1. It categorizes all variables as according to its variation in the dataset. In this method, we got different plotted lines. Each line in the plotted area represents principle component importance of the mentioned variable according to its variance. The dataset consisted of eleven variables. The PC1 showed 24% of total variance in the first component. That shows its importance and significant impact on the total data set. That component showed that it distinguished behavior because of its greater variation among the data set, and for this reason, we are taking this PC into account. On the graph, the eight principle component showed more than 85% of the variation in the total data set. Thus, it reduced the dimension of the data by eliminating three other variables. Those three variables showed a minor variation in the dataset. Therefore, they have been discarded.

Now, the component that was derived on the graph has Coefficient values. Coefficient value has a significant importance in the formation of the effect of principle component Fig. 1. The PCA generalizes the plotted graph from higher to smaller as according to coefficient value. This value indicates the impact of the specific trait on the outcome variable. The study was to examine the factor that had close relation and dependent behavior to the yield.

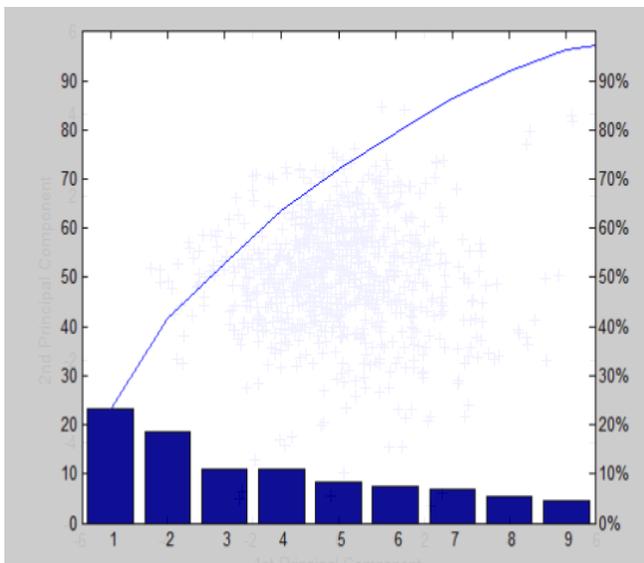


Fig. 1. PCA under normal water conditions.

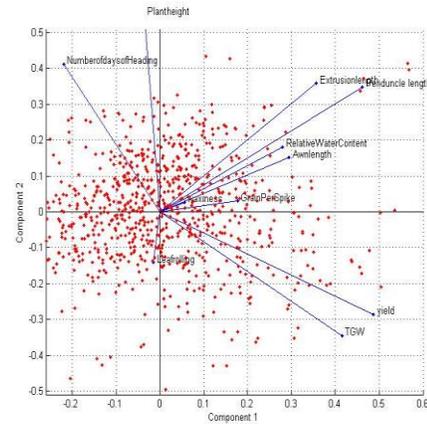


Fig. 2. PCA under normal water conditions using 2Dimensional graph.

The graph in Fig. 2 shows PC1 along x-axis and PC2 along the y-axis. From the graph, it can be visualized that trait “yield” has higher coefficient value among all other variables. So in 1st principle component “yield” has higher influence that shows that under the normal condition yield has higher relation and can be obtained at its maximum level. Whereas in the 2nd component along the y-axis, the coefficient value of “Plant height” is greater as compared to all other variables. Its participation in the variation of the 2nd component is higher. In the comparison of the PCA 3 and PCA 4, it was noticed that the variation in PC3 was because of the higher value of “waxiness” that was noticed 0.67 and the variation in PC4 were because of “RWC” that was 0.56 (Fig. 1). Other principal components also showed their coefficient values on the graph. So in the normal condition of the water, yield plant height and other variables have stronger and dependent relation.

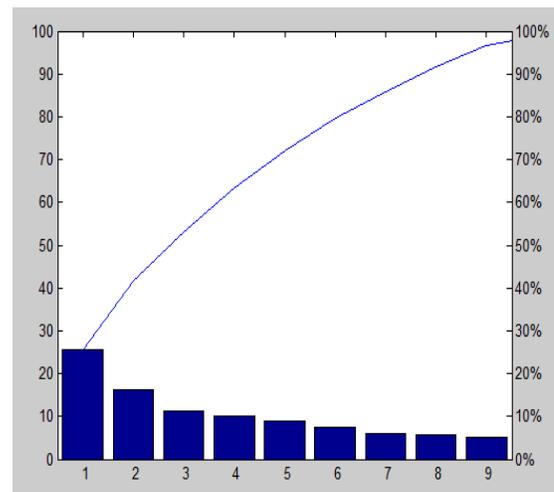


Fig. 3. Variation of data under stress water conditions.

In the second experimental phase, we provided eleven variables to the principal component analysis method as shown in Fig. 3. These were collected under water stress condition. It categorizes all variables according to its variation in the dataset. It has the same number of data set but was noted under stress condition of the water. Here the first

principle component that has the higher variation among the all data set, showed 25% of the total variation among all. That shows its overall impact on data. If we observe the graph in Fig. 3, it can be seen that the “pendulence length” has higher variation among all other traits.

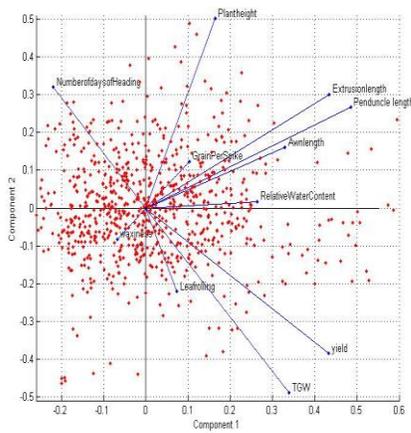


Fig. 4. PCA under stress water conditions using 2Dimensional.

So in the first principle component, the involvement of the “pendulence length” is more effective as compared to other traits or variables. The second trait that shows its dependency on the graph was “yield”. In 2nd principal component that is along the y-axis, the variable “plant height” has greater coefficient value. In stress data of the principal component 3 and PC4, the “grain yield” and “plant height“ values were observed at the highest level among all other variables. We got 9 principle components out of 11. Each has different variations and impacts. These components showed 90% of the total variation. It means that they are effective in the specific condition where the data has been collected in Fig. 4.

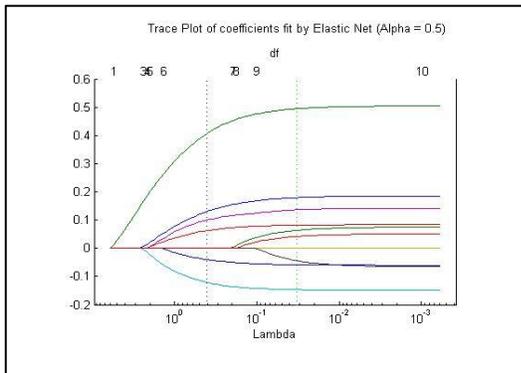


Fig. 5. Elastic net shows wheat traits variation in normal water conditions.

In lasso and elastic net, the expression of lambda is used. We used cross-validation to find the optimal value of regularization parameter λ with both minimum MSE shown in Fig. 5. Here the total values are plotted on the graph. Each line on the graph has separated color and represents the separate value. The dashed vertical lines represent the Lambda value with minimal mean squared error (on the right). Here when we applied data to the lasso technique. The first line of the graph was of TGW. That trait is one from our large data set. As the

lambda value against “tgw” is decreasing the MSE value is increasing. The smaller the predictor lambda, the greater is the MSE value; hence the dependency among variables will be less shown in Fig. 6.

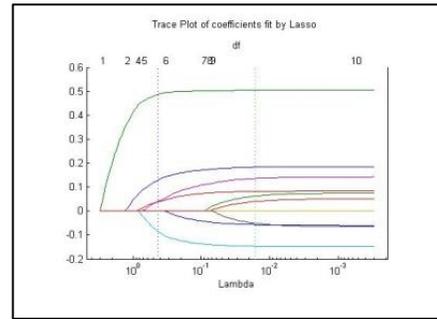


Fig. 6. Lasso shows wheat traits variation in normal water conditions.

The next variable is “grain per spike” its MSE is increased towards the right side of the graph as lambda is decreasing. So the point where it has larger none zero MSE values can be the best value of lambda to reduce the dimension of data and for finding the dependency of variables with yield. Where their lambda is between 10^0 and 10^{-1} at that value, there exist a minimum of the non zero values of MSE. So the “tgw”, “g/spk”, “rwc”, “heading”, “endurance length” are in a strong dependent relation with the yield under normal water condition shown in Fig. 6.

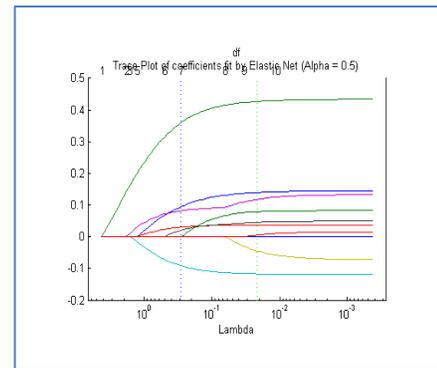


Fig. 7. Elastic net shows wheat traits variation in stress water conditions.

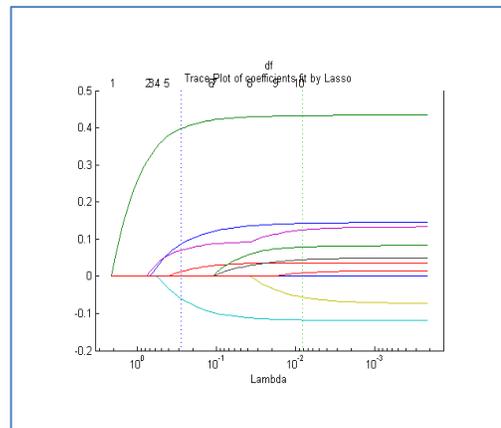


Fig. 8. Lasso shows wheat traits variation in stress water conditions.

Here, when we applied our data on the lasso and elastic net under stress condition, there produced a few results shown in Fig. 7 and 8. This result was based on the MSE values that showed the dependency of the variables to the yield. The values under stressed condition showed the different result in the lasso elastic graph. It was observed that the “TGW”, “pendulous length”, “grain per spike”, “heading” and “relative water content” had the higher values of the mean square when the lambda was declined. So a strong relationship between these variables and yield under stressed condition was observed. It means these variables were affected by the stressed condition as shown in Fig. 7 and 8.

A. Data Modeling in Normal water condition

In RBNN, we trained the neural network under the data set of normal water condition. Here, a single-layered architecture was used. This model consists of 100 numbers of neurons in hidden layer and has one output layer which is conventionally contained a single neuron.

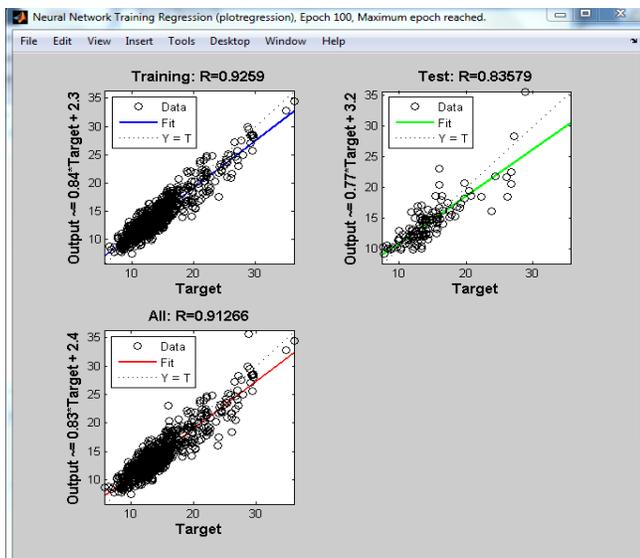


Fig. 9. Regression graph under normal water conditions.

Here radial basis function (RBF) was used as the activation function. A number of epochs in that model was 100. In radial basis neural network, when we use “trainlm” or “trainbr” activation function, the result obtained from this experiment is shown as in Fig. 9. Here the network is trained using the data set of those values which were observed under normal water conditions. Radial basis network works according to a total number of epochs unless the error is 0. Here is the plot of regression graph, the value of regression R was 0.91266. It was observed according to the radial basis function. The regression value indicates that traits and yield dependency was more than 90% and the yield relied completely on water.

The water factor showed significant impact on the wheat crop production i.e. the more water we provide, the more production of wheat we achieve. Just as yield was affected, different traits were also affected the same way in proportion to the supply of water as shown in Fig. 9.

TABLE I. RADIAL BASIS NEURAL NETWORK FOR WATER NORMAL CONDITION OBSERVATIONS

	Pr#	No of neuron	Activation Function	Training Function	R
Normal Condition	1	40	RBF	Trainlm	0.894
	2	40	RBF	Trainbr	0.904
	3	100	RBF	Trainlm	0.823
	4	100	RBF	Trainbr	0.912

We also applied some different techniques and methods in this neural network model to have different Regression values as shown in Table I. It helps to find the best relation between the traits and yield because of water. It also shows neural network architecture model is the best model to determine such observations. Different training functions were used for training dataset. In the first experiment namely activation function and “trainlm” training function, 40 neurons were used along with “RBF” model. By this experiment, we got 0.894 regression value. Then we trained the network model with the same architecture, and the output as regression value was 0.904. In next experiment, we used 100 neurons with activation function “RBF” and training function “trainlm”. The regression result we got was 0.823 Table I showing above.

B. Data Modeling in Stress water Condition

TABLE II. RADIAL BASIS NEURAL NETWORK FOR WATER STRESS CONDITION OBSERVATIONS

	Pr#	No of neuron	Activation Function	Training Function	R
Stress Condition	5	40	RBF	Trainlm	0.876
	6	40	RBF	Trainbr	0.899
	7	100	RBF	Trainlm	0.856
	8	100	RBF	Trainbr	0.896

In radial basis neural network model for the measurement of yield under stress water condition, we also implemented some more techniques. Those techniques consisted of the use of a different number of neurons and different training and activation functions [11].

First, we used 40 neurons and RBF as activation function, “trainlm” used as training data. In this, the result we achieved, was 0.876 regression value. Then with the same architecture, we did little change and used “trainbr” as training function. This time the regression value we achieved was 0.899. In the next experiment, the number of neurons was increased up to 100 and “trainlm” was used for training showing in Table II. The R regression value was 0.856. Same as on the similar architecture we applied “trainbr” training function and got 0.896 regression R value. All techniques and models presented different and identical values. It shows the dependency of yield and traits on the water Table II showing above.

VI. CONCLUSION

After applying different machine learning techniques on the field collected data that was observed, the growth ratio of wheat is found affected by water stress. Moreover, the wheat production has been reduced under the water stress condition. Many traits showed higher variation under stress condition. The variation shows that if there is lack of the availability of the water, then fewer yields will produce. Plant height and leaf rolling were affected under stress condition. Yield and TGW also showed significantly different ratio as compared to normal values. "Awnlength", "pendulacnelength", "extractionlength", "noofdaysheading" remain same under both conditions showing in Fig. 3. It means these four specific traits are survived under both water conditions. This factor showed significant impact on the wheat crop yield. It was also observed that the traits of the wheat were influenced by the water conditions showing in Fig. 4. Water condition showed significant effect on the growth and production of the wheat. The growth of traits, which influenced the yield, was reduced during that process. There was a significant relationship observed between different water condition and wheat and traits.

REFERENCES

- [1] P. R. Shewry, "Wheat," *Journal of experimental botany*, vol. 60, pp. 1537-1553, 2009
- [2] L. Akhtar, et al., "A review of hundred years of wheat research and development in Punjab (1911-2010)," *Pakistan Journal of Science*, vol. 62, pp. 128-134, 2010
- [3] F. Joint, et al., Evaluation of certain food additives and contaminants: seventy-seventh report of the Joint FAO/WHO Expert Committee on Food Additives, 2013
- [4] T. Umezawa, et al., "Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future," *Current opinion in biotechnology*, vol. 17, pp. 113-122, 2006.
- [5] R. Munns and M. Tester, "Mechanisms of salinity tolerance," *Annu. Rev. Plant Biol.*, vol. 59, pp. 651-681, 2008.
- [6] E. A. Bray, "Classification of genes differentially expressed during water deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data," *Annals of botany*, vol. 89, pp. 803-811, 2002.
- [7] M. S. S. Dahikar and S. V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach."
- [8] S. Emamgholizadeh, et al., "Seed yield prediction of sesame using artificial neural network," *European Journal of Agronomy*, vol. 68, pp. 89-96, 2015.
- [9] L. J. van der Maaten, et al., "Matlab toolbox for dimensionality reduction," *MICC*, Maastricht University, 2007.
- [10] O. Gonzalez-Recio, et al., "Machine learning methods and predictive ability metrics for genome-wide prediction of complex traits," *Livestock Science*, vol. 166, pp. 217-231, 2014
- [11] A. Galieni, et al., "Effects of nutrient deficiency and abiotic environmental stresses on yield, phenolic compounds and antiradical activity in lettuce (*Lactuca sativa* L.)," *Scientia Horticulturae*, vol. 187, pp. 93-101, 2015.
- [12] Adnan, Muhammad, M. Ahsan Latif, Abaid-ur-Rehman and Maria Nazir. "Estimating Evapotranspiration using Machine Learning Techniques. INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS, vol 8, no 9, pp. 108-113, 2017.