

Wheat Yield Prediction Using Neural Network and Integrated SVM-NN with Regression

¹ Naheed Ejaz, Shabbir Abbasi

Department of Computer Science

University of Agriculture Faisalabad, Faisalabad, Pakistan

Abstract- The production of wheat plays an important role in Pakistan's economy. Wheat yield forecasting is significant farming problem as it's the most important crop of Pakistan. Prediction of the wheat yield has been determined by data mining techniques with different environmental factors. Data mining techniques have been developed for analysing and implementation on wheat yield dataset to predict the yield which is very helpful to produce wheat. In this study, Neural Network and a Novel Integrated approach of Neural Network, Support Vector Machine and Regression are used to analyze and estimates the wheat yield production. We have used these predictive techniques with area, yield, production, soil pH, temperature, air pressure, rainfall, water availability, humidity, pesticides and fertilizer parameter for wheat yield prediction.

Keywords: Wheat yield, prediction, MLP Neural network, SVM, Regression, Integrated model

I. INTRODUCTION

Wheat, commonly called "Granum", has been the core source of food since ancient times. A considerable proportion of the world population depends on wheat as a sole method of earning. Apart from dissolving hunger, wheat stands as a vital protein and carbohydrate source. The abrupt arise in the wheat demand accounts for the hazardous climate alterations, including drought, flooding etc. The racing population concentration is responsible for the much-increased wheat and food demand. Therefore, the wheat crop reached its highest levels of momentous demand throughout the world. Evidently, engulfing 15% of the arability, wheat crop unveils its economic significance, as per records of 2011-2012. It was cultivated across 225 million hectors with the gross production of almost 704.1 million hectors tons [1]. The global wheat demand will be anticipated to beat 40% by 2030. To compensate this international wheat consumption soon, the cultivation rate would have been expected to increase up to 2 % per annum [2]. The raced wheat cultivation is the fundamental requirement keeping in view the demand for food supplies. Grain

¹naheedcheema27@gmail.com (Corresponding Author)

food, not is only a stuff of significant dietary elements, but also of imperatively defensive agents for human odd immune system. The lack in the supplements of such type cannot be patched up using some other kind of food stuff. So, the highly demanded wheat production has become a matter of national concern these days. The compromise on quality, preferring the quantity merely would originate allergic epidemics, maladies and incurable diseases, bearing further dilemma. [3].

A. *Economic Importance of Wheat*

Pakistan more importantly relies on its agricultural sector to bring out a committed livelihood for 70% of its population. Republic's 44 percent labour force is directly involved in agrarian field. Wheat being most reliable food source, so in Pakistan its cultivation is seen almost every corner of the country. Be-cause of increasing population rate, the productive demand of wheat has been considerably increased. Almost 3.1% increase in population yearly, the expected growth rate is smaller than the required rate[4].The gap diminishing among manufacture and demand of wheat yield is the main challenge for the scientist. As the agricultural land is to be decreased that cause the decrement in production quantity of wheat yield and fetching tough to achieve the demand, the gap between production and demand could be condensed by increasing the land for farming purpose. An enormous number of wheat could be formed with the addition in terrestrial area. The increase in cultivation area appears to be impossible due to increasing population, weather changes due to global warming, water pressure, temperature and some other features. However, we might upsurge the manufacture of wheat by refining the procedures and through undertaking well management in this field and usage of land, aquatic, and other environment changes variables concluded to achieve huge quantity of wheat. This invention could be a principal to improve development in the yield production. [5].

The economy of Pakistan depends upon the agriculture which subsidizes 26 percent to the GDP (gross domestic product) and it has great impacts on social welfare, employment and foreign exchange incomes. In agricultural crop, wheat add 40 % value, 28% cotton, 15% rice, 10% value added by sugarcane. In the production of wheat yield Pakistan have 10th and 59th ranked between the wheat producing nations with respect to area and yield respectively.

The estimation of future wheat yield has a great consequence but yield forecasting is very problematic task. Every farmer has desired to attain good yield and the prediction of yield can have helped to farmer to improved crops production. The soil type, humidity, area, production, surface temperature, pesticides, water availability and precipitation taken as constraint in

forecasting the crop yield in the agriculture field. Water irrigation system, ancient agricultural method, the quantity and quality of seed have effect on crop yield [6]. By using historical data, we can identify the crop losses so farmer can be prohibited from these losses by look after the crops. By using data mining methods, we can obtain more precise result in the prediction of crop yield [7].

B. Crop Yield Prediction in Pakistan

Pakistan is wheat producing country but not meet the desired demand so to fulfill the demand, every year Pakistan make the wheat import policies so it's very contemptuous whereas Pakistan is an agrarian state with productive and fruitful land. The Administration of state can take preventive techniques to eliminate the reasons and know those features that severely disturbing the crop production by early estimating the yield of wheat crop. Pakistan produced the low yield of wheat because of using outdated and unproductive forecasting approaches.

The ministry of Food and Agriculture offered three consecutive approximations of wheat yield but due to some reasons these 3 assessments predicted wrong. The initial guess was submitted via revenues representatives based on the information of deputy commissioner that accumulated from tehsils revenues and this guess released on first February. The 2nd estimate released on first April that showed estimation for wheat production which review the previous approximation. The second estimation was based upon the area and yield per acre attributes and the reliability between these two-approximations confirmed by Agriculture statistical cell and agriculture department of Lahore. This constancy was con-firmed through earlier prediction of yield, precipitation and some supplementary features which affected the growth of wheat crop. However, the estimating yield of April is compared with the actual yield that released on 1st August and presented a huge difference. This variance in actual and estimating showing that the conservative technique is insufficient and unsatisfactory for the development of strategy making.

Producers, consultants, and agricultural related organizations have an interest in the topic of crop yield prediction. Crop yield is a system that embraced of complex communication between the crop related features that required a model that predict the future yield before the production of crop. Yield predict-ing attitude is built on the numerous types of information which collected from diverse source such as atmospheric facts, Agro-meteorological like phenology and yield, water holding capacity and remotely sensed, agricultural statistics. Estimate of any crop before

harvesting is an important job to enhanced yield of crops and playing imperative role to make strategies intended for agronomic business [8], [9].

Data mining application mostly utilized for the estimation of crop yield. Data mining methods takes raw statistics and find out the useful information from that raw data. For discover this useful information, used data mining algorithms, arithmetical techniques, machine learning and database management systems. [10].

ANN, SVM with regression analysis has been assessed with examples that telling the how these techniques are applied to agricultural data sets. These procedures are used in estimation of crop production with the help of different environmental factors.

Our novel predictive model predicts the yield in harmony through the increasing estimation volume. This machine learning procedure has predicted that how much factors values affect the production of yield and association between them. It is regulating the regression of characteristics for wheat yield with the high accuracy rate. In this research neural network might suffered from over fitting and used computational knowledge through large number of neurons.

C. Motivation

In many articles, various data mining approaches and neural network techniques have been used to determine and gain more acceptable and better results within time. Neural network techniques are best method in yield prediction. In recent year integrated techniques have earned more attention by use of machinery leaning and computing. Integrated techniques are considered as best techniques to the de-termination then a single one. Is the integrated approach of support vector machine, artificial neural network with regression produces the same conclusive result for wheat yield within given time and less error rate is it so? well, in order to deal with this very matter, a forecasting model well be formed to assess and analyze that which prediction techniques give the most possible best results.

D. Objectives of this research

The main purpose of this vast research is the predictive determination of wheat using neural network and integrated approach of neural network, support vector machine with regression model. This study will mainly be based on various characters such as water provision, soil, production, rainfall, and temperature, which have a massive and direct effect on the crop yield.

This research will render best predicting and analyzing techniques to attain best possible results using Phyton and Weka tools.

II. RELATED WORK

AZhang and He[11]predict the yield of grains in 2001 and 2002 and take area effected by disaster, pesticides, irrigation system etc. as a self-determining variables. This study conducted on china country by using Back-Propagation NN and showed that forecasting accuracy 0.83% and 0.93% which represent a great development in prediction.

He et al. [12] proposed a parallelism technique and work as a nonlinear function and best way to ana-lyses the connection among complex attributes. In this work used ANN which is a type of that system. The output exposed that land related parameters which have direct link to yield of wheat that was water and nitrogen and the parameters that was multi-nominal to wheat yield ware potassium and some additional biological contents.

Crone et al[13], Explored the different estimation methods such as neural network and support vector regression and make comparison of these models. Result achieved from support vector regression and neural network showing that both gives relative precision and gives strongly high performance. Performance of support vector for regression and multi layered perceptron from artificial intelligence are investigated by compared with arithmetic prediction models.

Ruß et al.,[14]studied the useful techniques of modeling which used to uncover the already established patterns in the field of agriculture and using decision rule to transfer the raw dataset into precise statistic. Past dataset has great importance for prediction purpose. Data in the form of comprehensive picture and artificial neural network technique used to model this information.

Alvarez[15] made a model for crop yield prediction and to estimate productions. This analysis was con-ducted on wheat yield predictions in terms of environments parameters and soil physical appearance. Regression and ANN approaches were employed far statistics. This research showing that neural net-work technique foresees much better for wheat yield prediction in Argentine grassland.

Esfandiary et al.,[16]predict the wheat yield using weather defined attributes and agro related indices. This yields estimation done in 2 month earlier than cropping periods and represented that statistical techniques can accurately estimate the future yield. In this paper model gives the 83 percent precise prediction of wheat yield by using different indices of agriculture situation.

Marinkovic et al., [17] used data mining methods for yield estimating and showing that these approaches can straightforwardly implemented in agriculture field. Statistics set consist climate attribute, yearly production of diverse type of crops like soybean and maize that measured on regular bases. Results that achieved by data mining models compared with past techniques outcomes.

Cheng et al., [18] presented are practice of image identification to categorized of unreliable and ordinary types of wheat on the bases of exterior attributes. Attributes was studied by PCA algorithm. ANN also used in this study and created two layed using BP network method. 2.5% total Error was generated for wheat cataloguing.

Studied that neural network and SVM that gives outclass result in training and testing the dataset and make prediction. The result described by used of diagrams, accurateness proportion and numerical constraints. However, SVM is perform better in accuracy rate then neural network that are 78.6% and 83.50%.

Wang and Ma [19], said that, in china wheat stripe rust effected the wheat crop so forecasting of this disease is highly important to confirm the yield of wheat by governing and take management action timely. This prediction achieved by applying SVM and linear regression. This study represented, SVM can precisely predict the disease of wheat and information set used in better manner that related with the wheat streak decomposition. This model optimized the function and types of support vector machine for forecasting purpose. This model could be used in future on other type of plant.

Du et al., [20] presented wheat crop prediction using ground biomass approximation that needs essential satellite information for describe the plants position. The technique has instigated on winter wheat in Russia and outcome predicted from this algorithm almost near to actual values of wheat production that showed the high precision rate. Parameters such as irrigation, manure, environment variation and classes rule had significant consequence for wheat crop of winter and food security. In future this model can be improved using high quality pictures and atmospheric data collecting by observing field facts and remote sensing technique.

Romero et al., [21] the key determination of this research to predict the yield of durum wheat from different phenotype characters using diverse type of machine learning techniques under the weka tool and detected which one algorithms is the best by comparing all procedures. To achieve

this objective, implement One R, IBK, support vector machine and priori techniques. Outcome describe that the best algorithm was priori because it gives more accuracy for all environments.

Çakır et al., [22]proposed a model for prediction of yield by using ANN algorithms. Used climatic information like rainfall and temperature as a independent variable that effect wheat yield. Data consist of year 2012 to 2013 and 1st and 2nd year data used for trained neural network and for testing and forecasting the future yield of wheat respectively. In order to achieve the final results, linear regression techniques used numerous constraints for testing and dataset already has been used for ANN model. The consequences showed that multi linear perceptron technique would be able to achieve better performance as compared to regression technique and these estimation would be high if others attributes like fertilizer amount, time and status of crop are used during the development of system for prediction.

In this work procedural method proposed for prediction of wheat yield and extracted the features empirically and check the impact of stress conditions contains climate, ionic water and some additional constraints such as categories of soil, humidity, soil ph, N, K, C, Mn, S, deepness, high temperature, rainwater. In this work feed forward BP artificial neural network was used and concluded that weathers attributes, soil type and its configuration is interrelated with production frequency of crop yield. Neural network approach is a best model in the form of demonstrating and foreseeing is more actual to increase the amount of crop yield.

Kadir et al.,[23] studied MPL, BP based and feed forward ANN has been used to forecast the yield of wheat. Information obtained year wise from 1997 to 2007 that inculde environment related facts like sunshine, rainfall, frost, wind and temperature as input variables. This dataset separated into three files used for training, testing and validation purpose. After applying all types of neural network techniques it has been concluded that multi layered prepstone achieved high performance and shown 98% accurate results. However this neural network model applied on small dataset then accuracy rate would be low.

Guo and Xue [24] explains the neural network and spatial model for predicting the yield used in this research, output achieved by these approaches were according to precise mechanism. Ancient information of rainfall precisely used with the wheat yield, ANN and statistical techniques had been used for forecasting of wheat. Spatial-NN improved outcomes as associated to other neural network models with high accuracy.

Janjua et al., [25] analyzed the influenced of the weather situation on production of wheat crop and this research conducted by using of ARDL (Autoregressive Distributed Lag) method that evaluate the impression of weather on crops. Annual information collected from 1960 to 2009. Prediction expose that wheat crop does not affected by universal variations in the environment.

Stated that Multi-Layer Perceptron (MLP) back propagation-based- feed forward artificial neural net-work (ANN) used in prediction of wheat yield. Weather, soil condition, plant characteristics and crop management were taken as parameters. MLP was obtained 98% accurate result for prediction of wheat yield.

Ahamedet al.,[26] compose a study in which the main is purpose to emphasis upon the application methods of data mining such as linear-regression, clustering, K-means and K-Neural Network to guess the yield of foremost cereal yields by collecting the past information from the agricultural sector. A farmer can get a guess based on simple estimates made by this research about his plantation policy.

Geetha [27] make a investigation of different type of data mining models that are used in agriculture field. Techniques of data mining such as ANN, KNN, Decision tree, fuzzy set, Bayesian network, k mean and j48 are evaluated and assimilates the research of different scientists to in one system for the purpose of getting knowledge of all type of recent agriculture related application and techniques of artificial intelligence.

Kumar et al.,[28]build a crop selection method (CSM) to resolve the problem of crop selection and used meteorological conditions, soil type, water thickness, crop type as attributes for prediction model. By using this CSM, can be exploited the season crop by maximize the total yield in percentages and attained the maximum financial development of the nation. Crop slection method expanded high performance and accurateness that depend upon the on predicted values of influenced constraint.

Osman et al [29]proposed a model for prediction purpose and said that the influence of ecological conditions area of crop, production and crop yield is very important. This effort had been done to attain initial safety of food therefore in upcoming issues and circumstances could be easily fixed at maturity level of crop. For that purpose, machine learning systems were used. This technique could foresee good outcomes and identify category of crop with its yield, in initial phases of season crop 60 percent precision was measured and concluded that this approach gives better consequences as compared to previous approaches and considerable closer to past ones.

Presented a patrician model by integration of adjacent ultraviolet spectroscopy and SVM for modest, fast and precise analysis for the quality of wheat grains. PCA used to verify the possibility of categorizing the wheat data through near infrared methodologies after eradicate irregular information. Estimation values generated from support vector machine showing the precision frequency 100% and outcome attained from SNV are more precise for classification and identification of wheat quality so it can be used for real testing.

Shakoor et al., [30] In this paper studied the farming procedure to developed the valuable crop positions on the base of area that explored the crops that are most cost appropriate for agricultural in specific area of land. Six main crops of rice used to evaluate the Aus rice, Potato rice, jute and wheat yield and Aman and Boro rice crop. This estimation examined by administered machine learning models. Information collected from last years that taken from the Agricultural Research Council Bangladesh and Yearbook of Agricultural Statistics. In this study the ID3-Itrative Dichitomiser 3, KNN regression technique used and these models gives higher prediction results.

III. MATERIALS AND METHODS

Following algorithms has been used for yield prediction in this research; neural network and integrated approach of Support Vector Machine, Neural Network and Regression model. The results of the above-mentioned techniques have analyzed by applying statistical tool (Phyton/Weka). All the data is collected from concerned departments of agriculture and Statistic bureau of Pakistan which include last 21years record of wheat yield of Sargodha (district of Punjab. Water availability, area, yield production, rainfall, fertilizer, pesticides and maximum temperature would be the factors while collection of data.

A. Proposed Integrated Model

The planned model includes analysis the data firstly by these three algorithms using weka tool. Then make a new model for prediction of future wheat yield by combining neural network regressor, support vector machine regression and linear regression for best result using ensembling method.

B. Multi-Layer Perceptron NN

Neural networks are used almost in all field of life to crack numerous generic problems such as fore-casting, estimating, categorizing and cataloguing the data. The accuracy of these models is determined by the value of confidence level (R value) that is exist between 1 and 0. If value

closest to 1 it indicate the good accuracy and 0 value showing the worst consequences. Neural network has been applied on wide variety of area like services, health, design appreciation, relaxation, farming, authorization, indus-trial, disorder monitoring, language, script processing and biochemical investigation.

MLP (Multi-Layer Perceptron) is a supervised cataloguing system and is a most useful type of neural network as shown in figure 1. MPL model have capability of using the historical information for prediction and classification. In this technique data is alienated into 3 groups that are training, testing and validation set. Training set used to train the function and remaining set used for testing using past data. [31].

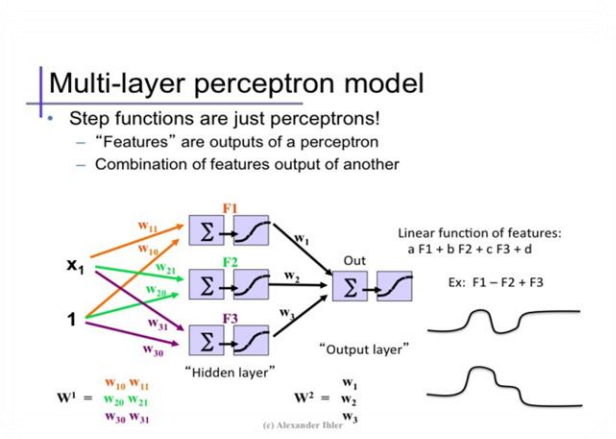


Figure 1: Generic Multi Layer Perceptron

To create such a network there is an input and output layer together with one or more hidden layers and each layer contain several nodes. Establishing the correct number of layers and nodes is problematic so, Python tool can assist in this area. We have applied neural networks multi-layer perceptron in my re-search. As we have data in excel file so, to convert that excel file into .csv file to make it practical in the neural networks. we gave that data to multi layered perceptron in first machine learning technique as following way:

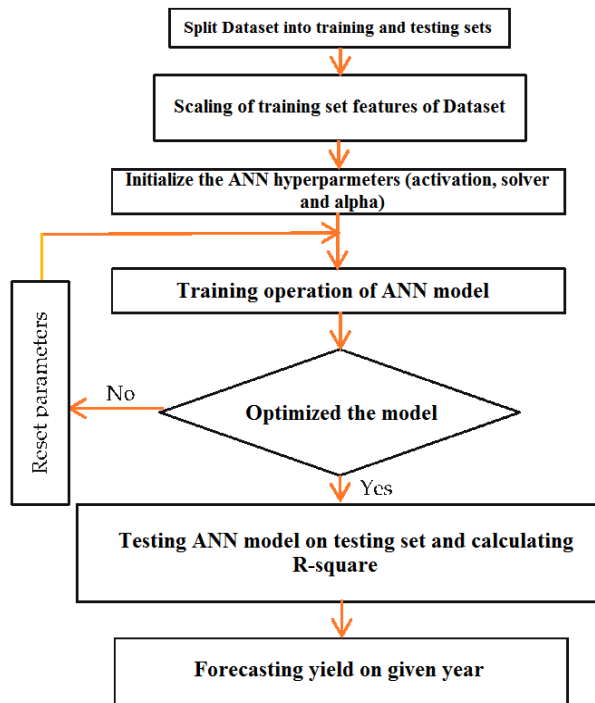


Figure 2: Flow chart of ANN

C. Linear Regression

Regression is an arithmetical model that used for measuring the association among the target variable and other self-governing attributes. In this research we have taken numerous characteristics as inputs such as rain, sunlight hours, moistness, soil ph etc. These characteristics are self-determining in nature so this model called as MLR (multiple linear regressions). Linear regression is used to create a model to create the connection among a scalar attribute and more than descriptive constraints by fitting a linear equation to the experimental statistics. [32].

Equation of Regression:

$$Y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$$

For this study, this equation is used to build the forecasting to achieved desired result. The equation transformed with the conformities of the study and takes the following shape:

- $Y = \beta_0 + \beta_1x_{area} + \beta_2x_{temp} + \beta_3x_{rainfall} + \beta_4x_{humidity}$

Where,

- Y is the yield
- β_0 is the y-intercept.
- β_1 is the coefficient between the area and the yield
- x area is the size, in hectares, of the area in that region
- β_2 is the coefficient between the temperature and the yield
- x temp is the value of the average temperature, in Celsius, in the region.
- β_3 is the coefficient between the rainfall and the yield
- x rainfall is the value of the rainfall, in mm, in the region.
- β_4 is the coefficient between the area and the yield.
- x humidity is the value of the humidity, in percentage, in the region.

D. Support Vector Machine Regression

When data is categorical then we used support vector machine. SVM produced a line to separate the dataset into classes, this process called as hyperplanes. This line makes a possible gap from these sets of information points as it cuts draw between them. The function that used for application of that line called as hyperplane equation and kernel function used to predict the target variable. The gap among points and the hyperplane take as error values but points accepted only when error rate is less then already defined rate of ϵ that treated as 0 value [33].

In addition, when data is numeric form and want to predict the target attribute then we can be used SVM for regression model. SVM regression has the goal to find out the function $f(x)$ that has a non conformity since the obtained target Yield value against the trained data. It can be explained as, when the errors are larger less than ϵ then we can't accept the error but the error has neglected in the case of less error then ϵ .

$$f(x) = (w, x) + b \text{ with } w \in X, b \in R$$

Where, (w, x) denotes the dot product in X. Flatness in the case of (3) means that one seeks small w. We find w and b by solving the following objective function using Quadratic Programming.

dataset applaid as following way:

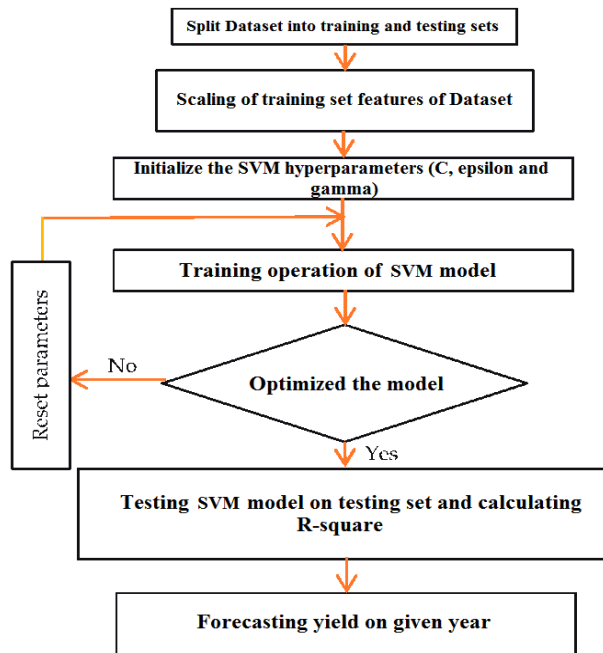


Figure 3: SVM Model

E. Dataset Used

All the data has collected from concerned departments of agriculture and Statistic bureau of Pakistan which include 1997 to 2017 record of one district of Punjab on the factors such as water, soil type, area, crop yield, production, precipitation, fertilizer, pesticides and maximum temperature.

- Temperature: Dissimilarity in temperature assuredly have an influence on the wheat production and the values receipts in Celsius degree.
- Area: For this research, yearly wheat production area measured in Hectares during the Kharif season (December to June) in Sargodha region of Punjab.
- Production: The wheat production values are collected in tonnes from the overhead cultured area during the yearly Kharif season (December to June) against Sargodha district of Punjab.

- Yield: Information collection against yield is depending on the wheat production and the cultivated area for wheat crop during Kharif season. This information measured in tonnes over hectares.
- Rainfall (mm): The average yearly rainfall is measured by calculating average from the monthly precipitation.
- Humidity: Like the way, a fact is collected for rainwater, I have also intended and got the average yearly moisture and values are takes in percentage.

F. Tools Used

Two data mining software are used in this study. At first, the dataset is entered in WEKA which analyzed by dividing the dataset in 8 folds and find out the predictions using SVM Regression, Multiple Linear Regression and MPL neural network. Phyton used is used to optimize the parameters of all these three models to gain highest accuracy after fit the dataset into it to calculate the predictions.

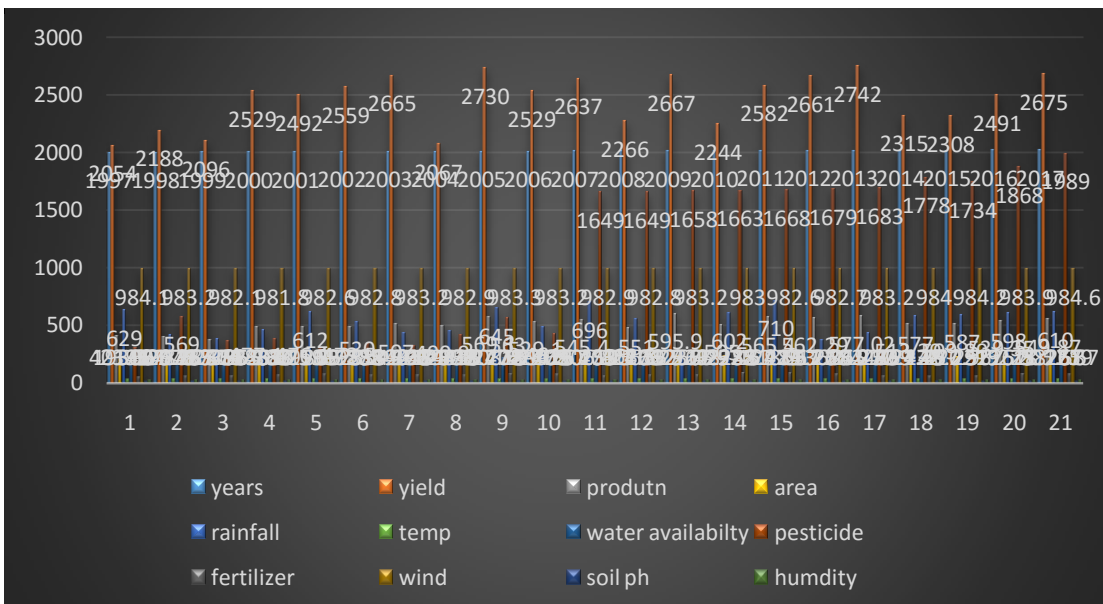
G. Optimization Strategy

In this strategy optimized the model by changing one parameter and keeping other constant. This procedure uses repeatedly for all parameters until to achieved maximum accurate results.

IV. EXPERIMENTAL RESULTS

The key purpose of this research is to explore the application of data mining to forecast the yield of wheat in Sargodha. To achieve this, last 21-year data is collected from the concerned departments of agriculture and statistical bureau of Pakistan so that an effective model can be build and tested.

Table 1: Dataset Graph



Our Dataset consists of 11 variables which features or input variables that are area, production, fertilizer, pesticides, humidity, rainfall, average temperature, water, soil ph and yield which is a target variable.

To analyze the dataset, we used Weka using neural network (Multilayer perceptron) algorithms for analysis. 8-fold cross validation method is used to divide the dataset for training and testing and choose following parameter

Table 2: Parameters used for ANN Model

Parameters	Optimal Value chosen
Learning Rate	0.3
Momentum	0.2
Seed	1
Hidden layers	25
Training Time	500

In this Algorithm we have used 25 hidden layers, learning rate 0.3. and gained 75.35 % accuracy. Support vector machine regression and linear regression achieved 0.8576% and 0.8656% respectively. This result showing that neural network gained less estimations compared to SVMR and Regression model. In order to to achieved more accurate results, implemented the proposed model and gained 0.839 % accurateness that presented, our integrated model has proved good predictive model then single neural network.

We have used MLPRegressor function from Scikit-Learn library of Python. MLPRegressor function implements Artificial Neural Network (ANN) which is a Multi-Layer Perceptron.

In this Artificial Neural Network, we have used 115 hidden layers. Activation function is set to Logistic while solver has been chosen to be Relu. we have divided my data into training and testing sets. The training set is 80% which is 16 out of 21 rows of data and testing set is 20% which is last 5 rows of 21 data.

Our Artificial Neural Network has been trained using maximum iteration of 100,000. Since, we have divided our data into training sets and testing sets, the training set gives us a training score of 1.0 that means model has been successfully trained. On the last five values of dataset of model has been tested, so, that an idea about performance metric on new data could be obtained. The five values on which trained models have been tested are 1997, 2017, 1998, 2008 and 2015 respectively.

Keeping all optimized parameters, it gives us a score (R^2) of 0.9037 on testing set. It gives a MEA for Artificial Neural Network of 77. 229370286. Also, RMSE of 81.86980727898153 for Artificial Neural Network (ANN). A graph has been plotted between Predicted Yield (Y-Axis) and Actual Yield (X-Axis) using ANN model, which provides the detail that both are linked with each other by showing this figure

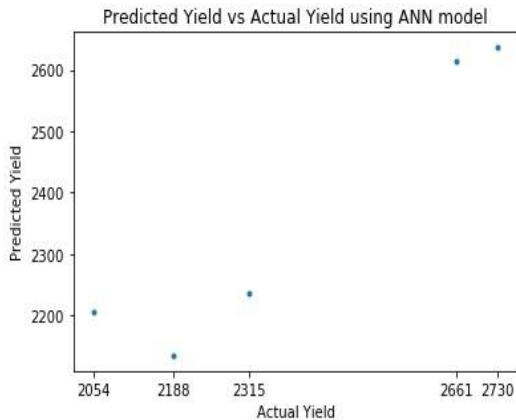


Figure 4: Predicted vs Actual Yield using ANN

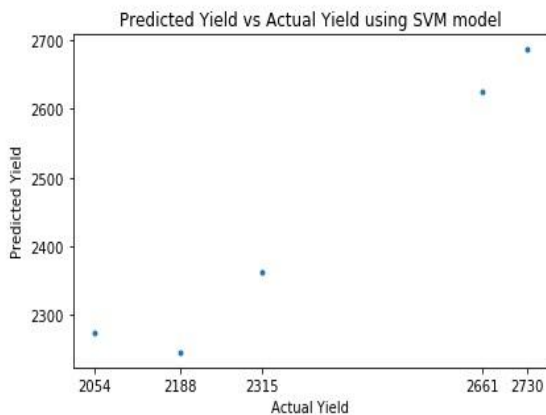


Figure 5: Predicted vs Actual Yield using SVM

Support Vector Regression (SVR) has been used considering cost parameter $C=2390$ that's a trade-off between performance and error. Cache size 1000, coef0 parameter 0.1, degree parameter 1, epsilon 0.0112999999 and gamma 0.0312999 has been used which is also shown in figure 5.

Table 3: Parameters used for SVMR Model

Parameters	Optimal Value
C	2390
Gamma	0.031299
Epsilon(ϵ)	0.0112999999
Cache size	1000
Kernel	Poly kernel

Keeping all these parameters, it gives us a score (R^2) of 0.8348 on testing set. It gives a MEA for Support Vector Machines (SVM) of 81.0976932529. Also, it gives a MSE of 11509.5250104 and RMSE of 107.2824543455161 for Support Vector Machines (SVM).

Linear Regression algorithm using Scikit-Learn library in Python has been carried out, with default parameters which are automatically specified by Scikit-Learn library. Linear Regression (LR) model gives following values of yield on the testing sets respectively 2166.79398916, 2226.55572033, 2690.93975094, 2211.57177945, 2752.32632769 against the year 1997, 2017, 1998, 2008 and 2015.

Keeping all these parameters, it gives us a score (R^2) of 0.93542593526 on testing set. It gives a MEA for Linear Regression of 55.4152253836. Also, it gives a MSE of 4499.07139557 and RMSE of 67.07511755911375 for Support Vector Machines.

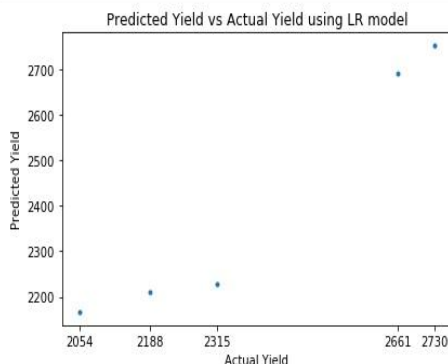


Figure 6: Predicted vs Actual Yield using Linear Regression

Whenever two approaches are used simultaneously for Regression, as Artificial neural network, Support vector machine and then Linear Regression are used here, Ensembling is required for them. In Regression based Ensembling, average outputs from ANN, SVM and LR respectively are used to get final prediction. The Ensembling technique allows to remove the trade-offs of all three approaches and ensures that the final prediction is not biased in any way.

Integrated model in phyton tool has been used in which the combined effect of every model could be obtained. We have used predictions value getting from neural network, support vector regression and linear regression to make an integrated model so, idea about performance metric on new data could be obtained. The five tested values of trained model are 1997, 2017, 1998, 2008 and 2015. The predicted yield for next year on the base upon given data using ANN, SVM and LR is: [2530.909]

Keeping all these parameters, it gives us a score (R^2) of 0.926362235855 on testing set. It gives a MEA for Integrated Model of 49. 0346255372. Also, it gives a MSE of 5130.5668868 and RMSE of 71.62797558775965 for Integrated Model. We have plotted a graph between Predicted Yield and Actual Yield using Integrated model. It gives us an insight on how the both are related to each other.

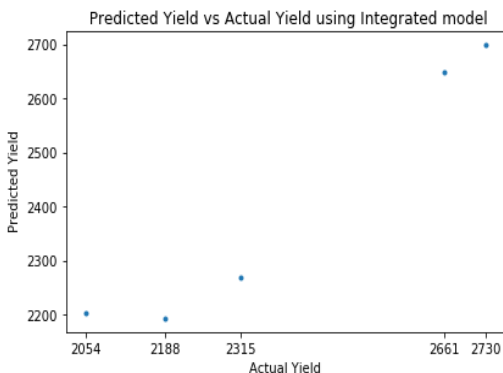


Figure 7: Predicted vs Actual Yield using Integrated Model

V. CONCLUSION

In the start of research some questions have been set. Results drawn made capable of giving answer to those questions. Answers have been provided in positive way machine techniques provide good opportunities to carry out accurate recommendations to measure the relationship

between yield and other variables, including attributes which are important with respect to yield and is integrated approach of support vector machine, artificial neural network and regression made a good productive model for yield prediction then neural network.

It was observed that integrated model gives maximum output as compare to neural network, we have also observed that support vector machine and regression can also be used for wheat yield prediction. Using all these three techniques in Weka the results have been analyzed by cross validation and this method gives the less results then the splitting test method. The performance of the neural network and integrated model analyzed by R-squared value. This study showed that the traits have significant effect on the wheat yield. All features have its individual importance. In this study, we have finally showed that an integrated model of neural network, support vector machine and regression model can be used for prediction of yield is emerging technology which provides reliable and fastest wireless communication service by using current spectrum band efficiently. System that uses fuzzy logic technique for spectrum management to make decision intelligently about spectrum sharing in CR network. In this paper we have presented about CR and dynamic spectrum access techniques and features. The proposed system is based on fuzzy logic. Fuzzy logic provides the efficient spectrum sensing so that existing spectrum can become more flexible.

REFERENCES

- [1] Kim, S., & B. E. Dale, 2004. Global potential bioethanol production from wasted crops and crop residues. *Biomass and bioenergy*, 26(4):361-375.
- [2] Dixon, R. A., and G. M. Pasinetti, 2010. Flavonoids and isoflavonoids: from plant biology to agriculture and neuroscience. *Plant Physiology*.154:453-457
- [3] P. Shewry, 2009. *Journal of experimental botany*. 60:1537-1553.
- [4] Akhtar, L., M. Hussain, A. Tariq, and M. Nasim, 2010. A review of hundred years of wheat research and development in Punjab. *Pak J Sci*. 40:128-34.
- [5] Okong'o, N., A. Leboissetier, and J. Bellan, 2008. Detailed characteristics of drop-laden mixing layers: Large eddy simulation predictions compared to direct numerical simulation. *Physics of Fluids*, 20(10):103305.
- [6] Ramesh, D., & B. V. Vardhan, 2013. Data mining techniques and applications to agricultural yield data. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(9): 3477-80.
- [7] Han, J., J. Pei, and M. Kamber, 2011. *Data mining: concepts and techniques*. Elsevier.
- [8] Supit, I., 1994. System description of the wofost 6.0 crop simulation model implemented in cgms. *Theory and algorithms*, 1: 146.
- [9] Stöckle, C.O., M. Donatelli and R. Nelson, 2003. Cropsyst, a cropping systems simulation model. *European journal of agronomy*, 18(3-4): 289-307.
- [10] Feelders, A., H. Daniels, and M. Holsheimer, 2000. Methodological and practical aspects of data mining. *Information & Management*, 37(5):271-281.
- [11] Zhang, Y. and Y. He, 2005. Study of prediction model on grey relational bp neural network based on rough set. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, (4764-4769).
- [12] He, Y., Y. Zhang, S. Zhang and H. Fang, 2006. Application of artificial neural network on relationship analysis between wheat yield and soil nutrients. *Engineering in Medicine and Biology Society, 27th Annual International Conference of the, IEEE*

- [13] Crone, S.F., S. Lessmann and S. Pietsch, 2006. Forecasting with computational intelligence-an evaluation of support vector regression and artificial neural networks for time series prediction. In Neural Networks, IJCNN'06. International Joint Conference on (3159-3166).
- [14] Ruß, G., R. Kruse, M. Schneider, and P. Wagner, 2008. Data mining with neural networks for wheat yield prediction. *Advances in Data Mining. Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, 47-56.
- [15] Alvarez, R. 2009. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *European Journal of Agronomy* 30(2): 70-77.
- [16] Esfandiary, F., G. Aghaie and A.D. Mehr, 2009. Wheat yield prediction through agro meteorological indices for ardebil district. *World Academy of Science, Engineering and Technology*, 49: 32-35.
- [17] Marinkovic, B., J. Crnobarac, S. Brdar, B. Antic, G. Jacimovic and V. Crnojevic, 2009. Data mining approach for predictive modeling of agricultural yield data. *Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment Novi Sad, Serbia*.
- [18] Cheng, F., F. Chen and Y. Ying, 2010. Image recognition of unsound wheat using artificial neural network. In *Intelligent Systems (GCIS), Second WRI Global Congress on* (172-175).
- [19] Wang, H. and Z. Ma, 2011. Prediction of wheat stripe rust based on support vector machine. In *Natural Computation (ICNC), Seventh International Conference on* (378-382).
- [20] Du, X., J. Meng, I. Savin and Q. Li, 2013. Wheat yield estimation in russia with modis time-series data based on light use efficiency model. In *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, 2848-2851.
- [21] Romero, J.R., P.F. Roncallo, P.C. Akkiraju, I. Ponzoni, V.C. Echenique and J.A. Carballido, 2013. Using classification algorithms for predicting durum wheat yield in the province of buenos aires. *Computers and electronics in agriculture*, 96: 173-179.
- [22] Çakır, Y., M. Kırıcı and E.O. Güneş, 2014. Yield prediction of wheat in south-east region of turkey by using artificial neural networks. In *Agro-geoinformatics (Agro-geoinformatics), Third International Conference on* (1-4).
- [23] Kadir MK, Ayob MZ, Miniappan N. Wheat yield prediction: Artificial neural network based ap-proach. In *Engineering Technology and Technopreneuship (ICE2T), 2014 4th International Con-ference on 2014 Aug 27* (pp. 161-165).
- [24] Guo, W. and H. Xue, 2014. Crop yield forecasting using artificial neural networks: a com-parison between spatial and temporal models. *Mathematical Problems in Engineering*.
- [25] Janjua, P.Z., G. Samad and N. Khan, 2014. Climate change and wheat production in pakistan: An autoregressive distributed lag approach. *NJAS-Wageningen Journal of Life Sciences*, 68: 13-19.
- [26] Ahamed, A.M.S., N.T. Mahmood, N. Hossain, M.T. Kabir, K. Das, F. Rahman and R.M. Rahman, 2015. Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in bangladesh. In *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 16th IEEE/ACIS International Conference on* (1-6).
- [27] Geetha, M., 2015. A survey on data mining techniques in agriculture. *International journal of inno-vative research in computer and communication engineering*, 3(2): 887-892.
- [28] Kumar, R., M. Singh, P. Kumar and J. Singh, 2015. Crop selection method to maximize crop yield rate using machine learning technique. In *Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), International Conference on* (138-145).
- [29] Osman, J., J. Inglada and J.-F. Dejoux. 2015. Assessment of a Markov logic model of crop rotations for early crop mapping. *Computers and Electronics in Agriculture*. 113:234- 243.
- [30] Shakoor, M.T., K. Rahman, S.N. Rayta and A. Chakrabarty, 2017. Agricultural production output prediction using supervised machine learning techniques. In *Next Generation Computing Appli-cations (NextComp), 1st International Conference on* (182-187).
- [31] Krikke, J. & Alfonsi, B., 2006. In the News. *IEEE Intelligent Systems*, 21(3):102–104, c3.
- [32] Ye, Nong, 2013; *Data Mining: Theories, Algorithms, and Examples*, CRC Press.
- [33] Dey, U.K., A.H. Masud and M.N. Uddin, 2017. Rice yield prediction model using data mining. In *Electrical, Computer and Communication Engineering (ECCE), International Conference on* (321-326).