ISSN:0976-0172

Dr.B. Nageshwarrao, JBio sci Tech, Vol 13(2),2025, 01-06

Journal of Bioscience And Technology
www.jbstonline.com

Using Artificial Intelligence Models To Estimate Wheat Yields And Comparing Them For Improved Prediction

[1] Dr.B. Nageshwarrao, [2] N. Supriya,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Article Info

Received: 30-04-2025 Revised: 16-06-2025 Accepted: 28-06-2025

Abstract—

As a new social issue, crop output prediction is a challenge that modern farmers must deal with. Focusing on two Artificial Intelligence (AI) models and their comparative analysis, this article aims to improve crop production prediction for famous wheat varieties. In order to forecast the wheat harvest, this article compares two AI algorithms. Estimating agricultural yields is a challenging task for farmers. The crofters will be able to anticipate the wheat crop production with the aid of our initiative. A loss of 11.1782 is produced by the artificial neural network, whereas a random forest approach yields a score of 0.999R2. The random forest technique outperforms the artificial neural network when it comes to forecasting.

Keywords—

Topics covered include agriculture, ANNs, random forests, and agricultural production prediction.

I. INTRODUCTION

The agricultural industry is crucial on a global scale. Agriculture is a multi-crop system. The three primary crops grown across the globe are maize, wheat, and rice. Some reasons, such as the extensive usage of pesticides, climate changes, unseasonable rainfall, etc., have diminished agriculture's prominence in modern times. People rely on agriculture as a foundation. One of the most widely cultivated crops worldwide is wheat. When it comes to wheat production, China is unrivaled. There will be 133,590 metric tons of wheat produced in China this year. Estimates for 2019 put global wheat output at 609,747 metric tons. In 89 different nations, 2.5 billion people consume wheat. The main goal of this study is to predict the future wheat crop production using historical data. The demand for food is growing at a quick pace these days due to the expanding population. Wheat yield information could be useful for farmers when tracking crop productivity. Although the land will remain the same, the people will grow in the days to come. There will be less food shortages in the future as crop yields improve. We are using machine learning and deep learning methods to do regression analysis in our project. A combination of an artificial neural network and a random forest algorithm was used to forecast the wheat harvest. Our models were fed winter wheat data.

II. Methodology

In order to forecast the wheat harvest, the suggested techniques include random forest and artificial neural networks. While comparing the outcomes of equally effective ANNs RF (random forest) and... We have been given seven aspects of winter wheat crop data as input data. In the end, we get an accurate yield prediction from our regression models.

ISSN:0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

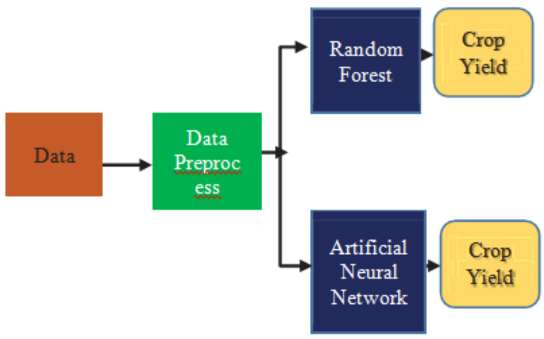


Fig.1.proceduretodevelopmodelsforwheatyieldpredictio n.

III. Random Forest Regression

As a supervised learning method, Random Forest takes training data in two forms: input and output combinations. Both regression and classification tasks are handled by a random forest method. It's a method for bagging.

Random forest = Decision tress + row sampling with replacement + column sampling + aggregation.

Decision trees are prone to overfitting as the depth grows. We have used decision trees as base learners in random forest, however there is no suitable depth for these trees. The training process becomes much easier with a big dataset that has a decent quantity of features. Reducing the model's variance is the primary benefit of the random forest. Categorical characteristics with several categories will not be well-suited for random forest. We developed a random forest system to forecast wheat harvest yield in our experiment. Ten decision trees have been used as first learners. The final result is the weighted average of all the predictions made by the base learners.

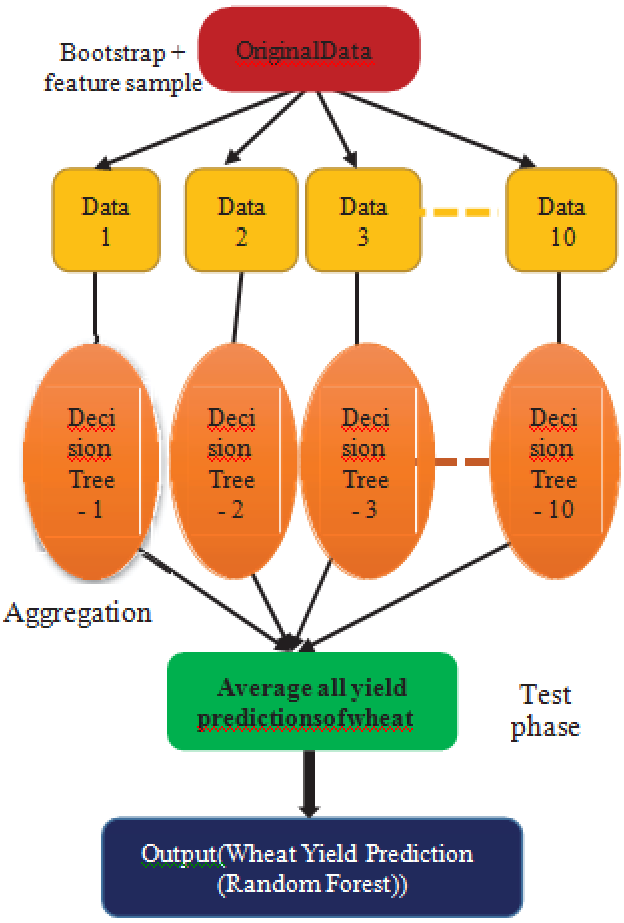


Fig.2flowchartforrandomforestregressionalgorithm(wheatyiel dprediction)

a) **R2 Score:** It is the ratio of explained variation in y to the total variation in y.

a) **HYPERPARAMETERS:** The number of base learners (decision trees) is an important hyper parameter in a random forest. We are used ten estimators (base learners or decision trees) in our model.

IV.Model Design for Random Forest

ISSN:0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

V. ARTIFICIAL NEURAL NETWORK

A neural network is another name for an artificial neural network, or ANN. Brain networks were the primary source of inspiration. For our regression challenge, we used an ANN, which accepts input data and processes it to produce output. For ANNs (artificial neural networks) to be able to forecast output. It performs well when the training data is substantial; otherwise, it fails miserably. For this specific issue, we are modeled as a neural network for the purpose of regression analysis. With the same number of neurons as input qualities, this basic model contains two completely linked hidden layers. As an activation function for buried layers, we are using reluas. For the output layer, no activation was applied. An optimization approach called SGD was used.

VI. Model Design for the Artificial Neural Network

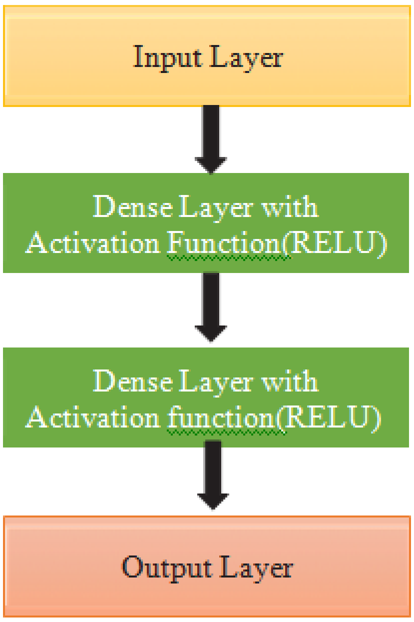


Fig.3.Architectureforpredictionofcropyieldusingneuralnetwork.

Scatter Gradient Descent (SGD) It quickly determines the update by calculating the derivative from each training set. B. Loss Function a) Mean Squared Error (MSE): It's a tool for regression analysis. Finding the average of the squared difference between the expected and actual values is what it is.

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

A. Metrics:

MAE (*Mean absolute error*): It is used for regression. it is the mean of the absolute difference between the actual values and predicted values.

$$MAE=\frac{1}{n}\sum_{j=1}^{n}|y_j-\hat{y}_j|$$

b) Mean Squared Logarithmic Error (MSLE): This is the percentage difference between the actual and projected values.

$$MSLE=\frac{1}{n}\sum_{i=1}^{n}((\log y_i+1)-(\log \hat{y}_i+1))^2$$

c) MAPE (mean absolute percentage error): MAPE is abbreviated as MAPD. By dividing the actual value by the anticipated value, we can get the variance between the two. Each anticipated point in time has its absolute value added to it, then divided by the number of fitted points (n), and finally multiplied by 100.

$$MAPE=\frac{\sum\frac{|A-F|}{A}\times 100}{N}$$

d) Hyperparameters: A crucial hyperparameter in neural networks is the Hidden layer. In our model, we have used a mere two hidden layers. 500 was the epoch value chosen.

VII. IMPLEMENTATION

The data will be divided as follows: 70% for training, 10% for validation, and 20% for testing. Both a random forest technique and an artificial neural network were used to address the situation at hand. The random forest method takes 22 seconds to train, and we have utilized the R2 score as a statistic. We used relu as the activation function in our artificial neural network, with mean squared error as the loss function and SGD as the optimization algorithm. It takes about 25 minutes to train the neural network model. 500 was the epoch value chosen.

VIII. DATA

The data was retrieved from Github. The collection includes wheat crop statistics from many US countries. Every day, they are supplying the winter heat data. The values in the dataset

ISSN:0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

are separated by commas. The data set has 182549 rows and 26 columns, representing the number of characteristics. We have not used all of the columns in our model in our experiment. We have employed seven characteristics and removed certain columns from our data file based on their relevance. Our model uses the following seven characteristics as input (x) and output (y) in its wheat crop prediction: latitude, longitude, apparent temperature maximum, apparent temperature minimum, temperature maximum, temperature minimum, and humidity.

IX. DataPre-Processing

Due to the removal of some columns, not all of the columns in the dataset have been used. In addition, we used the minmaxscaler on the data. Using minmaxscaler, any outliers in the data may be removed.

X.RESULTS

For this experiment, we employed two methods to forecast wheat harvest yields: a random forest algorithm and an artificial neural network. A loss of 11.1782 is produced by the artificial neural network model, whereas an R2 score of 0.9999 is produced by the random forest technique. It takes 22 seconds to train the random forest algorithm. It takes 25 minutes (500 epochs) to train the ANN model.

A. RESULTFORRANDOMFORESTALGORITHM

Model	Number of base learners	R_2scoreon test data
Random forest algorithm(used for regression)	Ten	0.999993836 2663

Table 1. shows the results for wheat yield prediction using random forestalgorithm.

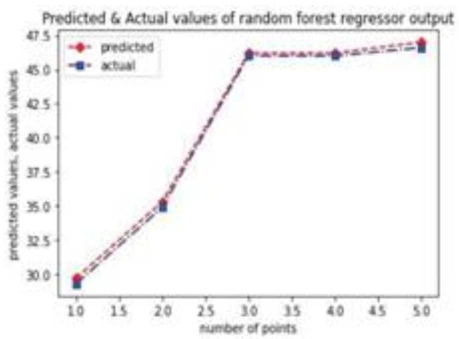


Fig4:Graphforpredictedvaluesandactualvaluesusingrandomforestalgorithm

B. ResultsforANN:

Model	Number of hidden layers	Loss(Mean square error)	Metric(Mean absolute error)	Validation loss (Mean square error)	Validation for metric(Mean absolute error)
ANN	Two	11.1782	2.4065	12.8319	2.5641

Table 2. shows the results for wheat yield prediction using artificial neuralnetwork.

Model	Numberof hidden layers	metric(Mean square logarithmic error)	Metric(Mean absolute percentage error)	Validation (Mean square logarithmic error)	Validation for metric(Mean absolute percentage error)
ANN	Two	0.0127	8.2961	0.0127	9.3929

Table3.showstheresultsforwheatyieldpredictionusingartificial neuralnetwork by taking different performance metrics.

C. Graphs:

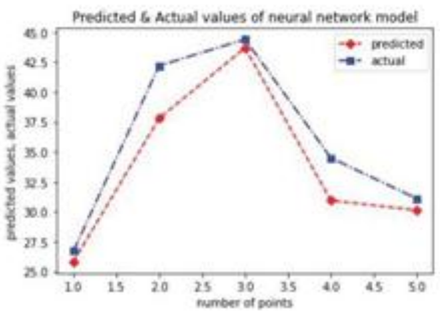
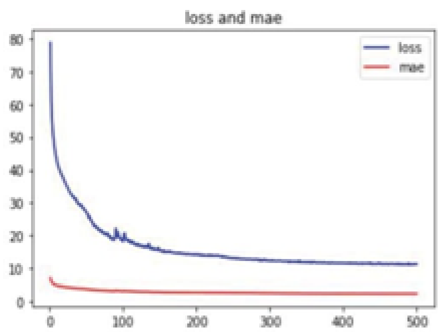


Fig5.GraphforpredictedandactualvaluesusingANNmodel



ISSN:0976-0172

Journal of Bioscience And Technology
www.jbstonline.com

Fig6.Graphformeansquareerror(loss):

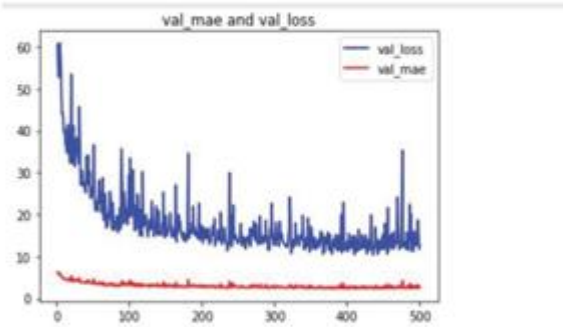


Fig7.Graphformeanabsoluteerrorfortrainingandvalidation MAE.

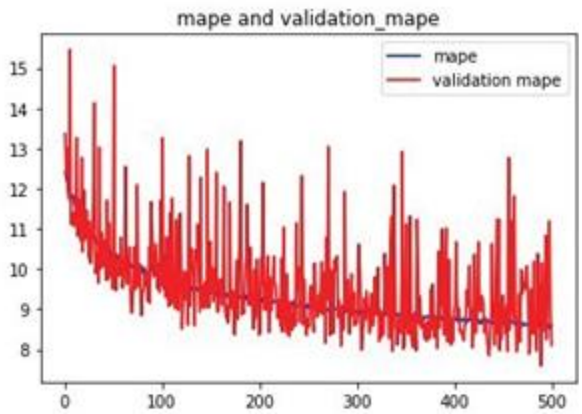


Fig8.Graphformeanabsolutepercentageerrorfortrainingandv alidation MAPE

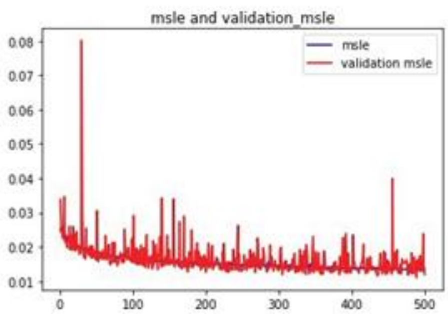


Fig9.Graphformeansquarelogarithmicerrorfortrainingandvalid ationMSLE

XI.Conclusion

The outcomes of the two methods are being compared. The methods used include artificial neural networks and random forest regression. Both are functioning well. A.N. networks aren't very effective, but the Random forest method is. However, many farmers still struggle with finding reliable methods for estimating crop yields. The most accurate forecasting of wheat output is something that this study will assist farmers with.

XII.References

[1] YanghuiKang,Mutiu Ozdogan, Xiaojin Zhu,Zhiwei Ye,ChristopherHain and Martha Anderson, "*comparative assessment ofenvironmental variables and machine learning algorithms for maizeyield prediction in us mid west*", environmental research letters,published in 19 may 2020.

[2] NariKim,Kyung-jaha,No-WookPark,JaeilCho,SungwookHongandYang-Won Lee,"*a comparison between major artificial intelligencemodels for crop yield prediction: case study of the midwestern unitedstates"*,2006-2015,internationaljournalofgeo-information,publishedin 21 may 2019.

[3] Trung-Tin Tran,Jae-Won Choi,Thien-Tu huynh Le and Jong-wookKim,a comparitive study of deep CNN in forecasting and classifyingthemacronutrientdeficienciesondevelopmentoftomatoplant,MDPI,published in 17 april 2019.

[4] M.Kumar,A.Bandyopadhyay,N.S.Raghuwanshi,R.Singh,comparitivestudyofconventionaland artificialneuralnetwork-basedETo estimation models , springer- verlag 2008 published in 20 may 2008.

[5] Perpetua Noronha,Divya.J,Sruthi.B.S,comparative study of datamining techniques in crop yield prediction, international journal ofadvanced research in computer and communication technology,October16.

[6] HetalPatelandDharmendraPatel,acomparative studyonvariousdatamining algorithms with special reference to crop yieldprediction,Indian journal of science and technology,June 2016

[7] D Ramesh,B Vishnu Vardhan, analysis of crop yield prediction usingdataminingtechniques,internationaljournalofresearchinengineeringand technology,Jan 2015.

[8] Mr.Vikas Lamba,dr.v.S.Dhaka,wheat yield prediction using artificialneuralnetworkandcroppredictiontechniques(asurvey),internationaljournal for research in applied science and engineeringtechnology,September2014.

[9] Prashanth Govardhan, Rasika korde,Rashi lanjewar,survey on cropyield predictionusing data miningtechniques,International journalofadvance in computer and electronics technology,December2018.

[10] Dr.A.Senthil Kumar,P.Arun, a survey on agriculture analysis for cropyield prediction using data mining techniques, IOSR journal ofcomputerengineering(IOSR-JCE).

[11] E.Manjula,S.Djodiltachoumy,a model for prediction of crop yield,international journal of computational intelligence andinformatics,march2017.

[12] Andrew crane droesch machine learning methods for crop yieldprediction and climatic change impact assessment in agriculture,environmental research letters,October 2018

[13] DivyaBH,ManjulaR,SivaBharathiS,MadhumathiR,asurveyoncropyield prediction based on agricultural data, international journal ofinnovative research in science engineering and technology.

[14] ArunKumar,NaveenKumarVishalVats,efficientcropyieldpredictionusing machine learning algorithms, international research journal ofengineering and technology,June 2018