



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 6, June 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Efficient Model to Predict Crop Yield using ANN

Prof. Archana Kale¹, Prof. Gajanan Aochar², Rushikesh Pawar³, Sanika Thorat⁴, Himani Bankar⁵,
Abhishek Bagul⁶

Assistant Professor, Department of Computer Engineering, MES College of Engineering, Savitribai Phule
Pune University, Pune, India^{1,2}

UG Students, Department of Computer Engineering, MES College of Engineering, Savitribai Phule Pune
University, Pune, India^{3,4,5,6}

ABSTRACT: One of the most pressing needs of the hour is a paradigm for crop yield forecast. Droughts and other crop failures have become more common, resulting in significant loss of life and livelihood around the world. The growing occurrences of global warming and shifting weather have resulted in higher crop output uncertainty. The agricultural production has also been difficult to discern due to the uncertainty. This warrants a need for a crop yield prediction approach which has been achieved in this proposed research paper. Therefore, to enhance the process of crop yield prediction, this research analyzes the applications of machine learning approaches. The presented approach in this research utilizes Linear regression along with Artificial Neural Networks and Fuzzy classification. The approach has been evaluated for the purpose of presence of any error which has resulted in satisfactory results of implementation.

KEYWORDS: Linear Regression, Artificial Neural Network, and Fuzzy Classification.

I. INTRODUCTION

India is an agrarian country. An agrarian country is a country that is primarily reliant on the cultivation of crops and other agricultural products. Feeding a country is a herculean feat which is achieved effectively through cultivation. India has the largest areas of fertile land that can be used for cultivation. Due to the country being an agrarian country, majority of the individuals are dependent on the agriculture for their livelihood. Crop Cultivation is a major profession in the Indian subcontinent that supports a large number of individuals. There are extensive number of professions that rely on the crop cultivation indirectly for their effective execution and implementation.

The crops are grown across the world are highly dependent on a variety of attributes and parameters. For successful crop cultivation and an effective result is needed for the purpose of achieving maximum economic stability. As failed crops lead to a loss of economy that can be highly crucial for the survival of the largely agrarian economy. Therefore, conditions of drought can be a death blow to the country that is entirely reliant on the crop cultivation and agrarian activities. Therefore, there is a need for an effective technique to identify the crop yield in advance to prepare for the imminent arrival of the drought or favorable conditions.

The various attributes that govern the crop yield are highly complex and require in-depth understanding of the paradigm of agriculture to provide any type of prediction. The weather conditions over a course of time and the effective utilization of the available data can hold the key towards an accurate as well as an effective mechanism for prediction. The predictions can be highly useful in premeditation of the events that could be catastrophic for the farmers in a large agrarian economy.

There have been several researches that have been performed for the purpose of enabling effective predictions on the crop yield for a certain time period in the future. These researches have been studied in detail for the purpose of understanding the different approaches that the researchers have employed to achieve the yield prediction. These researches have been useful in achieving an effective prediction that can be useful for providing a valuable insight towards the crop prediction effectively.

The researches that have been performed are analyzed thoroughly to understand the usefulness and the various limitations that are prescribed by the authors. The researches have been identified to achieve an effective implementation of the crop yield prediction with some minor discrepancies. These discrepancies have led to the reduction in the effective accuracy of the system that is unacceptable in a prediction environment. Therefore, this research has helped achieve the goals for the development of an effective crop yield prediction methodology based on the Artificial Neural Networks and Fuzzy classification approaches.

The proposed technique is deeply narrated in the section 3. The experimental evaluation is performed in section 4 and whereas section 5 concludes this research article with the scope for future enhancement.

II. LITERATURE REVIEW

A. Manjula analyzes farm management in terms of input and outputs besides preserving resources is carried out by the precision agriculture technique. Towards precision agriculture, data mining techniques are used most of the time for beneficial results. [1] To build various indices for assessing the productivity of crops numerous efforts have been made to exploit remote sensing data. The indices such as Temperature Condition Index (TCI), Vegetation Condition Index (VCI), and Normalized Difference Vegetation Index (NDVI) are included. In the proposed paper they have used a flexible and extensible technique called has extensible Crop Yield Prediction Framework (XCYPF). For crop yield prediction for rice and sugarcane crops rainfall data and surface temperature with the available indices.

A. Terliksiz suggests to have sufficient crop production is very necessary in the coming days because the world population is constantly increasing. For the economic development of a nation monitoring crop growth and yield, estimation is very important, it directly knocks on national and international economies and plays important role in food management and food security. [2] In the proposed paper the researcher used deep neural network models such as Convolutional Neural Network (CNN) and Long-Short Term Memory Network (LSTM). The proposed study focuses on the soybean yield prediction of Lauderdale County.

Y. Gadge specifies the major source of living for the people in India is agriculture and agriculture-related industries it is a major source of economy of the country. [3] The proposed paper used various data mining techniques used for predicting the crop yield. Data mining techniques are mainly divided into two-part Classification, Clustering. Data mining procedure is separated into seven methods Data cleaning, Data integration, Data selection, Data transformation, Data Processing Pattern estimation, Knowledge display. The results obtained by various algorithms are summarized by various authors for crop yield prediction.

M. Paul narrates for agriculture information technology has become more and more a part of our day to day life. In large datasets, unknown and likely impressive patterns were discovered by the data mining. [4] Many algorithms and other techniques are created to extract knowledge from large sets of data. For classifying unknown samples using the information provided by a set of classified samples different types of classification techniques. [4] In the proposed paper the soil dataset is taken from the soil testing laboratory Jabalpur, M.P, and two important and well-known classification algorithms K-Nearest Neighbor (KNN) and Naive Bayes (NB).

A. Nigam claims agriculture is one of the major and the least paid occupation in India. [5] By applying various machine learning techniques paper focuses on predicting the yield of the crop. By considering factors like temperature, rainfall, area, etc. will help farmers to decide which crop to grow to get the maximum yield by using. India ranks second in this sector of agriculture. India's GDP is steadily declining with the country's broad-based economic growth due to the revolution in industrialization. To perform accurate prediction and handle inconsistent trends machine learning algorithms like RNN, LSTM is used. Thus the prediction made by machine learning algorithms will help the farmers

M. Liu describes the ridge-furrow tillage as one seedbed design practice widely used for soil and water conservation in agriculture which is formed by soil cultivation. In the proposed technique oilseed rape is used it is a plant widely planted in rain-fed croplands of the semiarid area.[6] They measured soil water content, aboveground biomass of oilseed rape at its non-identical growth stages. Most of the researches are based on a comparison between conventional field tillage and Ridge-furrow tillage with the same configuration and seldom referred to the influence of surface configuration on soil water content and plant growth under ridge-furrow tillage. Thus it an effective cultivation technique for harvesting water and promoting the growth of individual oilseed rape.

Meeradevi states the required data from a large dataset is extracted from data mining. Data mining contains two types of techniques such as clustering and classification. [7] Data is classified into predefined classes or groups due to the lack of technology and scientific farming methods. Clustering divides the data into similar groups and hence can be treated as one section to perform the analysis. To train the dataset ARIMA model is used. On individual factors such as location, farm size, temperature, rainfall, and various crop dataset is optimize and then the detailed recommendation is provided to farmers. The prediction will also be based on the model type used to predict.

T. Islam elaborates agriculture is the essential ingredient to mankind which is a major source of livelihood. Due to its very fertile land and favorable weather, varieties of crop growth can be grown in Bangladesh so it is known as an agrarian country. For modeling and prediction, an artificial Neural Network is considered a robust tool. [8] To get better output and prediction the algorithm plays a major role. For comparing the accuracy and error rate support vector machine, Logistic Regression, and random forest algorithm are also considered in the proposed paper. Agricultural crop selection and yield prediction are done by the deep neural network. The machine learning algorithm can change our country’s agricultural image.

N. Gandhi aims at achieving high crop yields. To predicting crop yield under different cropping scenarios Artificial Intelligence (AI) techniques provide a major effective approach in the last decade. In the proposed model researchers describes the development of a rice crop yield prediction model through the use of ANNs. [9] The parameter discussed in the study is, minimum temperature, average temperature, maximum temperature, and reference crop evapotranspiration, area, production, and yield for the Kharif season. The data that is extracted is from the WEKA tool using Knowledge Flow.

N. Gandhi expresses in India rice crop production provides more than 40% of overall crop production. Researchers provide a decision support system prototype for rice crop yield prediction for Maharashtra state, India. [10] The proposed paper predict crop productivity under different climatic scenarios this can assist farmers and other stakeholders in making important decisions in terms of agronomy and crop choice. The Graphical User Interface (GUI) allows for the selection of the range of precipitation, minimum temperature, average temperature, maximum temperature, and reference crop evapotranspiration and predicts the class of yield viz., low, moderate, or high.

J. Ren presents for national food security and sustainable development of society. In China, winter wheat is one of the most important food crops. Early and accurately getting the knowledge of crop yield or changes of yield can help farmers. Remote sensing, agro-climate model, crop growth model, and statistical sampling methods are some techniques and methods to estimate crop yield. In the proposed paper to calculate PAR and used MODIS data to calculate fPAR they have used global TOMS ultraviolet reflectance data

Y. Zhang explains for guiding agricultural production timely and accurate knowledge of crop yield information is of great significance. To detect the photosynthetic capacity indexes yield formation, spectral analysis technology, as a fast and non-destructive method. The neural network prediction model is optimized by the genetic algorithm. Employing the calculating spectral vegetation index or spectral shape parameters grain yield was estimated. Yield estimation accuracy could achieve good results thus the advantage was simple and easy and in particular conditions. Thus the proposed paper achieves good results offered rapid, convenient, and valuable guidance for the agricultural production

III PROPOSED METHODOLOGY

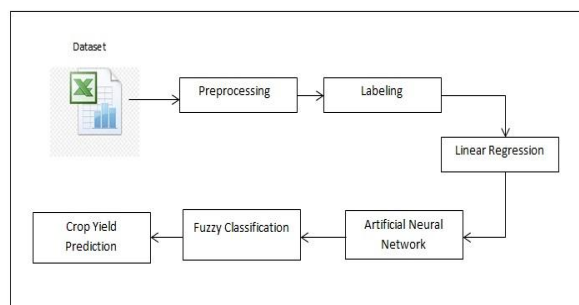


Figure 1: Proposed model System Overview

The proposed methodology for the purpose of enabling a crop yield prediction approach through the use of Artificial Neural Networks has been depicted in the system overview diagram given in figure 1 above. The prediction system has been achieved through the use of a procedure whose steps have been detailed below.

Step 1: Dataset collection, preprocessing and Labeling –The prediction system for the crop yield has been demonstrated through the use of a dataset downloaded from the URL: <https://github.com/VaibhavSaini19/Crop-Yield-Prediction-using-ML>. This dataset is useful in determining the yield of the crop to provide as an input to the proposed methodology.

The extracted dataset is provided as an input to the proposed methodology for performing the predictions. The dataset is interfaced in the form of a workbook that is read by the system through the use of the JXL API. This Library provides the java code the ability to interface with the workbook file.

The dataset contains a large number of attributes out of which only certain attributes are under consideration for our implementation that can be useful for attaining the precise prediction. The extracted attributes are the area of the field, yield, temperature, precipitation, humidity, soil type, crop type, and district. These are the relevant attributes and the rest of the attributes are irrelevant and they are eliminated. These attributes are useful in determining the prediction and has been effective in realizing the improvement in the yield which can be highly useful.

The Graphical user interface is utilized to register the user before providing it an input to the system in the form of the users attributes. These attributes include the use of name, DOB, sex, email addresses, mobile number, User ID and password. These attributes are the used to perform authentication to gain access to the proposed system. This system can now be used to provide the parameters of the current field for the purpose of current yield estimation. These values are provided to the system for Linear regression evaluation.

Step 2: Linear Regression –In this step of the procedure the previous step provides the input in the form of a double dimension list. This list consists of the selected attributes that are preprocessed and labeled. The regression on these attributes is achieved through the use of the linear regression approach.

The linear regression achieves a regression between an independent and a dependent variable that changes constantly. These values are referred to as the x [] and a y [] list, where x is the independent value and y is the dependent value respectively. This is illustrated mathematically by the equation 1 given below.

$$Y = Mx + B \quad (1)$$

This equation is utilized to attain the value of regression by the value of b which is the intercept withm which is the slope and is unknown. These values are achieved by providing the dependent and independent values, which are the attributes such as temperature, precipitation and humidity. These values are provided from the input preprocessed and labeled dataset in the form of an array X [] for intercept and slope measurement through the equations 2 and 3 given below. And the values that are provided for the same attributes through the user interface is considered as Y [] list to feed into the mentioned below equations.

$$M = \frac{N \sum(xy) - \sum x \sum y}{N \sum(x^2) - (\sum x)^2} \quad (2)$$

$$B = \frac{\sum y - M \sum x}{N} \quad (3)$$

Where:

x = Independent variable (user input)

y = Dependent variable (Dataset values)

M = Slope or Gradient (how steep the line is)

B = the Y Intercept (where the line crosses the Y axis)

N = Size of the array

Y = Intercept value

The values once measured can be used in the equation 1 to achieve the value of the dependent variable. This is done by taking into account the user input from the system for a particular attribute. The values of the x or the independent variable are provided and the respective values of y are achieved as a regression. These values are utilized further for the process of crop yield estimation.

Step 3: Artificial Neural Network –This is the most important and crucial step for the methodology. This is the step that actually performs the estimation of the yield. The labeled and preprocessed list is provided as an input to this step of the approach for the hidden layer and the output layers evaluation. The attributes realized for achieving the linear regression previously, have been utilized as an input for this step of the methodology. This is useful as it allows for effective and accurate determination of the crop yield.

The preprocessed and labeled list containing the required attributes from the dataset is provided as an input to this module for the evaluation of the hidden and output layers. The Artificial Neural Networks execute in the form of layers which are, namely, input layers, hidden layers and output layers. For the prediction of the crop yield the input values of the eight attributes, the temperature, precipitation, area, yield, humidity, soil type, and crop for a particular district is provided as an input.

These values are used along with the provided random weights for the purpose of generating the five hidden layers. The two weights for each of the attributes is used and combined with the bias weights. The five hidden layers are evaluated through the implementation of the Relu activation function. The attained hidden values are provided to the output layer for the evaluation of the output error probability.

The two output layer uses the five hidden layer values along with the weights and the bias weights to achieve the output layer values. These values are utilized in the equation 4 given below, along with the two target values, T1 as 0.01 and T2 as 0.99 to attain the error probability rate as given below.

$$\text{Error Probability} = \sum \frac{1}{2} (T_0 - O_L)^2 \text{----- (4)}$$

Where,

T = Target Values

O_L = Output Layer Values

This error probability value is then added at the end of the row in the double dimension list of the respective district. This is repeated for all the attributes. The resultant list is provided for sorting in the ascending order of the error probability rate, therefore, the topmost values are considered with the least amount of error. The error probability rate is inversely proportional to the accuracy of the probability attained which will be used in the next step for the classification procedure. The process of Hidden layer estimation can be depicted in algorithm 1.

ALGORITHM 1: Hidden Layer Estimation

//Input: Pre-processed ListPR_L, Weight set W_S= { }

//Output: Hidden Layer value list H_LV

hiddenLayerEstimation (PR_L, W_S),index=0

1: Start

2: H_LV = ∅ { Hidden Layer value}

3: **for** i=0 to size of PR_L

4: ROW = PR_L[i]

5: **for** j=0 to size of ROW

6: X=0

7: **for** k=0 to N [Number of Neurons]

8: ATR=ROW[j]

9: X = X + (ATR* W_S[index])

10: index++

11: **end for**

12: H_LV=reLUmax(0, X)

13: **end for**

14: **end for**



- 15: return H_{LV}
- 16: Stop

Step 3: Fuzzy Classification – The error probability list attained in the previous step is used as an input in this step of the process for classification through fuzzy classification. This list is already sorted in the ascending order, whose length is extracted and divided by 5. This value is used to segregate the list into 5 crisp values, with the varying fuzzy crisp values of VERY LOW, LOW, MED, HIGH and VERY HIGH. These groups are also ranked from 1-5 respectively.

The user input is also taken in consideration in this step of the procedure, and the corresponding district is searched in the clusters for its position according to the respective label. Once the position of the input region is achieved, it is divided by the total number of clusters which is 5, the obtained value is then combine with the regression value of the respective region achieved in the previous steps. This results in a range of different values that are then sorted for the purpose of classification and ratio evaluation. The best suitable outcome for the prediction is selected from the list to achieve the yield which is then displayed to the user through the interactive user interface.

IV. RESULT AND DISCUSIONS

The presented approach for the prediction of crop yield has been achieved through the use of Java programming language. The NetBeans IDE has been used for the development of the proposed technique. The machine for deployment is configured with 6GB of RAM and 1TB of Hard Disk which is powered by an Intel Core i3 processor. The JXL API is being utilized for enabling the interfacing of the dataset with the java code in a workbook format.

Experimental assessment has been achieved to extract an in-depth evaluation of the prediction technique for the incidence of any errors. The error assessment extracts the preciseness of the prediction that can be extremely fruitful in measuring the preciseness of the crop yield prediction system. The RMSE or the Root Mean Square Error performance evaluation technique has been utilized to attain the error of the presented prediction model.

The RMSE technique uses two correlated and continuous variables for the determination of the error between these two variables. The variables used in our methodology are the expected crop yield predictions and the obtained crop yield predictions. The equation used to calculate the error is shown in the equation 5 given below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_{1,i} - x_{2,i})^2}{n}} \quad \text{-- (5)}$$

Where,

\sum - Summation

$(x_1 - x_2)^2$ - Differences Squared for the summation in between the expected crop yield predictions and the obtained crop yield predictions

n - Number of samples or Trails

Table 1: Mean Square Error measurement

No. of Trials	No of Expected Crop Yield predictions	No of Obtained Crop Yield predictions	MSE
1	10	9	1
2	4	3	1
3	8	8	0
4	9	7	4
5	1	1	0
6	0	0	0
7	5	3	4
8	13	11	4
9	2	2	0
10	11	9	4

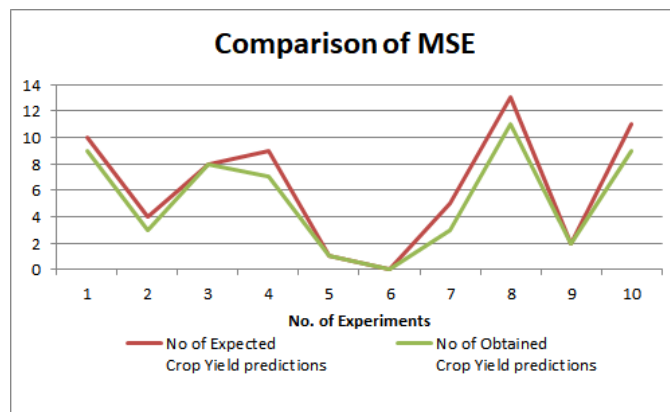


Figure 2: Comparison of MSE in between No of expected Crop Yield predictions V/s No of obtained Crop Yield rate predictions

The experimental procedure and its results have been shown in the table 1 given above. The values achieved in the table are used to plot a line graph as shown in the figure 2 given above. Through in-depth assessment of the graphically represented and the tabulated values, we can reach a conclusion that the error achieved in this procedure for crop yield prediction is minimal. A set of 10 experimental trails have been performed with varying input to achieve the calculation of MSE or Mean Square Error.

The experimental outcomes illustrate that the error attained in the prediction system is reasonable and acceptable. The error for the prediction is usually identified in the prediction models which perform predictions on real world data. There are a number of different scenarios that influence the crop yield predictions. The attained MSE and RMSE values of 1.8 and 1.34 respectively are extremely satisfactory and illustrate an accurate implementation of the crop yield prediction model.

V. CONCLUSION AND FUTURE SCOPE

The proposed methodology for the crop yield prediction approach through Artificial Neural Networks has been elaborated in this research article. This approach utilizes a dataset and the user attributes as an input into the system. The dataset and the user input contain relevant information about the crops. This data is first preprocessed to remove the irrelevant and the inconsistent data. This preprocessed data is provided to the next step of the procedure for labeling the data; the labeled data is then given to the next step of the procedure for evaluation of the regression. The linear regression is being used for the evaluation of the regression list which performs the regression and provides the output to the next section for neuron formation. The Artificial Neural Networks have been implemented using the Relu activation function to achieve the error probabilities in the attributes. These values are then used to achieve the output which is subject to the classification through the use of Fuzzy Classification to achieve crop yield prediction. The approach is evaluated for its performance which has resulted in a highly acceptable score.

The future research direction can be in the implementation of the crop yield prediction approach in the mobile application paradigm so that any mobile user can avail the service of Crop yield predictions.

REFERENCES

- [1] C. Martinez-Felix et al, "Agricultural Monitoring, an Automatic Procedure for Crop Mapping and Yield Estimation: The Great Rift Valley of Kenya Case" IEEE LATIN AMERICA TRANSACTIONS, VOL. 17, NO. 10, OCTOBER 2019.
- [2] Jie Sun et al, "Multilevel Deep Learning Network for County-Level Corn Yield Estimation in the U.S. Corn Belt" IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 13, 2020.
- [3] Laura Martínez-Ferrer et al, "Crop Yield Estimation and Interpretability with Gaussian Processes" IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 2020.
- [4] Pritam Bose et al, "Spiking Neural Networks for Crop Yield Estimation Based on Spatiotemporal Analysis of Image Time Series" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 2019



- [5] David B. Lobell et al, "Comparison of Earth Observing-1 ALI and Landsat ETM+ for Crop Identification and Yield Prediction in Mexico" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 41, NO. 6, JUNE 2017.
- [6] Hossein Aghighi et al, "Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI" IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, 2018. on Computational Intelligence and Communication Networks 2015.
- [7] Michele Meroni et al, "Evaluating NDVI Data Continuity Between SPOT-VEGETATION and PROBA-V Missions for Operational Yield Forecasting in North African Countries" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 2018.
- [8] Zhihong Zhuo et al, "Regional Grain Yield Response to Climate Change in China: A Statistical Modelling Approach", IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 7, NO. 11, NOVEMBER 2014.
- [9] Roberto Luciani, "Agricultural Monitoring, an Automatic Procedure for Crop Mapping and Yield Estimation: The Great Rift Valley of Kenya Case", 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics) 2019.
- [10] Dhivya Elavarasan et al, Sandhya, "Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications", IEEE Access, 2020.



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details