Machine Learning Approach for Crop Yield Prediction: A Review

1. Archana Kale
2. Gajanan Aochar

1. Rushikesh Pawar
2. Sanika Thorat
3. Abhishek Bagul
4. Himani Bankar

Modern Education Society's College of Engineering,
19, Late Principal V. K. Joag Path,
Wadia College Campus, Pune-411001
Computer Department
Savitribai Phule Pune University

***Abstract* — Agriculture is one of the earliest and the most important sets of tasks that have been performed to achieve nutrition for human beings. Growing crops allowed humans to progress towards settling down in a specific location and form civilizations. Agriculture still plays a major role in fulfilling the need for a food source. Agriculture is also one of the most vital aspects of the Indian peninsula as it provides valuable income to the majority of Indian citizens. This is the reason why there is a critical need for monitoring the crop yield. To achieve effective prediction in the crop yield, there have been several types of research, but most of them have been analyzed to achieve a quite low accuracy. The low accuracy can be attributed to the highly complex paradigm for the prediction of crop yield which is dependent on the various attributes such as weather conditions etc. Therefore, to enhance the process of crop yield prediction, this research analyzes the applications of machine learning approaches. The analysis has been effective in achieving the methodology which employs the use of Linear Regression and Artificial Neural Networks along with Fuzzy Classification to achieve effective crop yield prediction.**

Keywords:*Artificial Neural Networks, Linear Regression 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, the majority of the individuals are dependent on agriculture for their livelihood. Crop Cultivation is a major profession in the Indian subcontinent that supports a large number of individuals. There are an extensive number of professions that rely on the crop cultivation indirectly for their effective execution and implementation.

The crops 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.

This literature survey paper dedicates section 2 for analysis of past work as a literature survey, and finally, section 3 concludes the paper with traces of future enhancement.

1. II RELATED WORKS

A. Manjula analyzes management of the farms through the use of input and outputs along with preservation of the resources carried out by the precision agriculture technique. Towards precision agriculture, data mining techniques are used most of the time for beneficial results. [1] The assessment procedure has built a collection of indices that determine the productivity of the crops through the use of the sensing data collected remotely. The assessment is based on indices such as Normalized Difference Vegetation Index, Temperature Condition Index (TCI), and Vegetation Condition Index. In the proposed paper they have used a flexible and extensible technique called has extensible Crop Yield Prediction Framework. The rainfall data and surface temperature with the available indices are utilized by the authors to estimate the crop yield for sugarcane and rice crops.

A. Terliksiz suggests that sufficient crop production is very necessary in the coming days because the world population is constantly increasing. For the development of a nation economically, the monitoring of crop yield and growth, and their estimation is very important, it directly knocks on the economy nationally and internationally along with playing a vital role in food security and management. [2] In the proposed paper the researcher used deep neural network models such as Long-Short Term Memory Network or LSTM and Convolutional Neural Network or CNN. The proposed study focuses on the prediction of soybean crop yield in Lauderdale County.

Y. Gandge specifies that for the residents of India the major source of income is crop cultivation and cultivation-related occupations; it is the biggest source of economy for the country. India also suffers from regular disasters such as flood or drought which causes a lot of damage to the crop. [3] The proposed paper used various data mining techniques used for predicting the crop yield. Data mining approaches are segregated into Classification tasks, and Clustering tasks. Data mining process is further divided into the following steps: cleaning of the data, integration of the data, selection of the data, transformation of the data, Estimation of the data mining patterns, Knowledge display. The results extracted by a collection of algorithms which are summarized by the researchers for the prediction of crop yield.

M. Paul narrates for agriculture information technology has become an integral part of every one's life. In large datasets, the insightful and outstanding patterns were discovered through data mining. [4] Many algorithms and other techniques are defined to extricate knowledge through huge datasets. The classification of the samples that are not known are utilized for the information extracted through the samples that are already classified using different types of classification techniques. [4] The authors have utilized the naïve Bayes algorithm along with the K Nearest Neighbor for the classification purpose on the soil dataset provided by the Jabalpur Laboratory for soil testing.

A. Nigam claims cultivation of crops is one of the biggest occupations and it is also the lowest paid profession in India. [5] Through the application of different machine learning approaches this publication concentrates on the crop yield prediction. By considering factors like temperature, rainfall, area, etc. can assist the farmers in their decision to choose the best crop for maximum yield. In the agriculture sector the Indian subcontinent ranks second. Due to growth and industrialization the GDP of India is reducing significantly. Machine learning algorithms such

as RNN, LSTM provide accurate and highly precise predictions. The machine learning algorithms assist the farmers greatly through the predictions.

M. Liu describes the tillage using ridge-furrow as one used for soil and water saving for crop cultivation through a seedbed design approach commonly used for soil based cultivation. In the proposed technique oilseed rape is used it is a plant widely cultivated in rain-fed farms of the semi dry area.[6] The authors calculated the biomass aboveground, soil water content of oilseed rape at its non-identical growth stages. Most of the research is based on a comparison between conventional field tillage and Ridge-furrow tillage with the same settings and very rarely affects the surface setting which is based on the water content of the soil and growth of the plant through ridge-furrow tillage. Thus it an highly useful agriculture approach for saving water and promoting the growth of oilseed rape plant.

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 because of the absence of scientific and technologically sound farming techniques. Clustering process segregating the data into equivalent forms this makes the data uniform for performing the analytical procedure. To train the dataset ARIMA model is used. On individual factors such as various crop dataset, rainfall, temperature, farm size, location, and is optimized and then the detailed recommendation is provided to farmers. The prediction can be based on the type of the model used to perform the prediction.

T. Islam elaborates on the most useful approach to humans which is also biggest income provider. Due to the favorable weather and large patches of fertile land, a plethora of crops 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 effective prediction and output the algorithm plays a major role. For the purpose of comparison the error rate and precision of the random forest algorithm, Logistic Regression and support vector machine, are also contemplated in the proposed paper. Agricultural selection of the crops and prediction of yield are done by the deep neural network. The machine learning algorithm implementation has the potential to improve the agricultural image of our country.

N. Gandhi aims at achieving high crop yields. For the prediction of crop yield through the changing cropping scenarios the paradigm of Artificial Intelligence (AI) techniques provide a major effective approach in the last decade. In the proposed model researchers utilized the ANN platform to achieve an effective and useful crop prediction approach. [9] The parameter discussed in the study is, yield for the Kharif season, production, area, reference crop evapotranspiration, maximum temperature, average temperature, minimum temperature, etc. The data that is extracted is from the WEKA tool using Knowledge Flow.

N. Gandhi expresses that in India rice crop production constitutes 40% of the entire production of crops. Researchers utilize the area of Maharashtra for defining a decision support system for crop yield prediction. [10] The proposed paper utilizes different weather situations for the prediction of crop yield. This can be helpful for farmers for taking valuable and informed decisions to increase crop yield. The Graphical User Interface (GUI) can be used to select the range of evapotranspiration, maximum

temperature, average temperature, minimum temperature, etc.

J. Ren presents for societal development sustainably and provides food security nationally. In China, the most important crop for food is the winter wheat. Early and accurately getting the information related to crop yield can help farmers. Statistical sampling methods, crop growth model, agro-climate model, and remote sensing are some approaches to predict the crop yield. In the proposed paper the crop yield has been predicted effectively using an effective and useful approach.

Y. Zhang explains that guiding agricultural production accurate and precise information about crop yield related information is extremely important. To identify the capacity of the photo sensitivity indexes along with the spectral analysis and yield formation is regarded as a rapid and non-invasive technique. The genetic algorithm is used to optimize the neural network based prediction approach. Employing the measurement of spectral shape related parameters or spectral vegetation indices the output grain yield was predicted. Yield estimation accuracy could achieve good results thus the advantage was ease of use in the specified conditions. Thus the proposed paper achieves good results offered fast, easy to use, and highly useful advice for the production of crops.

III PROPOSED SYSTEM

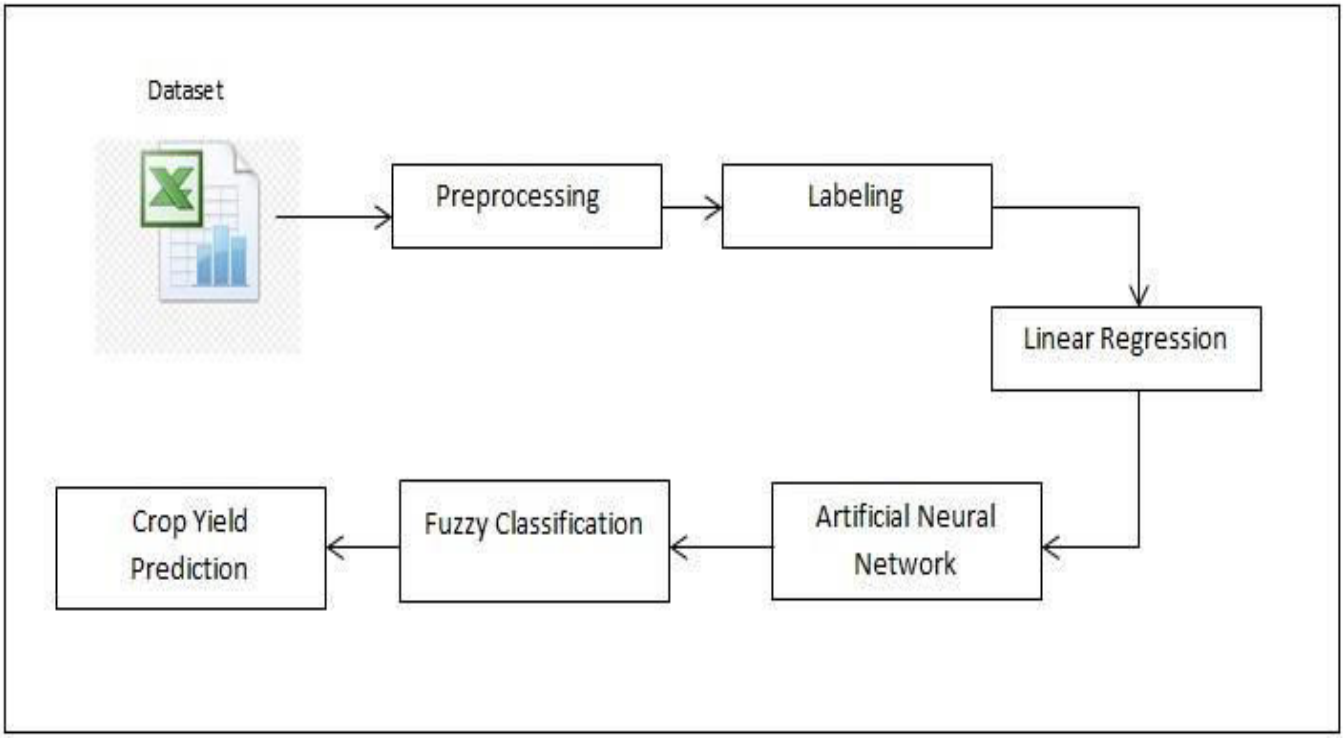


Figure 1: The Proposed model Overview

The proposed model for crop yield prediction is designed as depicted in the above Figure 1. The steps that are included in this process are broadly elaborated in the below mentioned steps.

Step 1: Preprocessing – As the spreadsheet dataset is fed to the system the complete dataset has been read in a double dimension list. This double dimension list is then subjected to select the most important attribute columns that eventually play the best role in identifying the crop yield. So for this purpose the proposed model selects the state of the soil and weather conditions, as they are more sensitive for the yield of the crop. These two attributes are collected in a double dimension list of n rows and two columns.

Step 2: Labelling – In this step the data is normalized by conversion into the equivalent integer format. As most of the attributes utilized for the crop yield prediction are in the string format. The string format cannot be effectively used for the purpose of enabling the execution of the system. Therefore, these string values are converted into integer equivalent forms. This is achieved through categorization of the attribute values and providing labels to the corresponding entries such as 1 for sunny weather, 2 for rain etc. This effectively labels the data and converts it into an integer format.

Step 3: Linear Regression – The Obtained preprocessed and labelled data is subjected to estimate the regression rates using the Linear Regression. Here the two columns of the state of the soil and weather condition values are stored in separate arrays like A and B. And these two arrays are used for the estimation of the gradient and y-intercept of the Linear Regression as m and b respectively. Then the obtained values of m and b are utilized for each of the instances of A array to call as x in the equation 1.

$$Y=mx+b \quad \text{_____ (1)}$$

Where
Y= Intercept
M= Gradient
B – Y Intercept
X – Instance of Weather condition.

Then the obtained Intercept value is measured for mid probability of the intercept values with each of the rows. Then the obtained rows are added into a list to call it as an ANN input list.

Step 4: Artificial Neural Network – The obtained ANN input list values are used to estimate the mean values of the soil state and weather condition values for each and every row. Then from these mean values the maximum and the minimum values are being calculated to term them as the target1 and target2 variables of the ANN.

After this process 10 random weights are being estimated in the range of 0 to 1 to assign different variables like W1,W2,W3,W4,W5,W6,W7,W8,B1 and B2. Here B1 and B2 represent the values of bias.

Using these weights along with the soil state and weather condition values hidden and Output layers are estimated using the equation 2 and Tanh activation function of ANN.

$$X1= AT1*W1 +AT2*W2 +B1 \quad \text{_____ (2)}$$

Where AT1 and AT2 are the attributes like Soil State, Weather Condition or Hidden Layer outputs.

The obtained output layer value is then measured for the minimum value to consider as the Probability value of ANN to form ANN probability List.

Step 5: Crop Yield Prediction through Fuzzy Logic – This is the last step of the proposed model, here based on the soil state and weather condition some crop yield prediction protocols are being classified in a list of Fuzzy crisp sets. This list contains some predictions like Very Low, Low, Medium, High and Very High.

This fuzzy crisp set is used to classify the ANN probability list in the inference Engine segment.

Once the classified values are being counted, then they are measured for their maximum value along with the attached crop yield level.

The obtained type of the crop yield level is displayed to the user through an interactive user interface.

IV MATHEMATICAL MODEL

(A)Set Theory

S= { } be as system for Crop Yield Prediction
Identify Input as S= {D_1, D_2, D_3….D_n}
Where L_F = Dataset
S= {D}
Identify C_{YP} as Output i.e. Mask Alert
S= {D, C_{YP}}
Identify Process P
S= {D, P, C_{YP}}
P= {P, R_{OI}, C_{NN}, D_T}
Where
P=Preprocessing
L_R =Linear Regression
A_{NN} = Artificial Neural Network
F_C = Fuzzy Classification
So the final System can be defined as
S = {D, P, L_R, A_{NN}, F_C, C_{YP}}

(B) SET DESCRIPTION:

2. Preprocessing
Set R_{OI}:
P_0 = Attribute List
P_1= Attribute Index
P_2= Attribute Selection
P_3= Selected Attributes List

2. Linear Regression

Set L_C:
L_{C0} = Selected Attributes
L_{C1}= Intercept Estimation
L_{C2}= Slope Estimation
L_{C3}= Regression List

3. ANN
Set A_N:

A_{N0}=Regression List
A_{N1}=Activation Function
A_{N2}=Hidden Layer Estimation
A_{N3}=Error Probability

4. Fuzzy Classification
Set F_C:
F_{C0}=Crisp values
F_{C1}=Fuzzifier
F_{C2}=Defuzzification
F_{C3}=If-then Rules

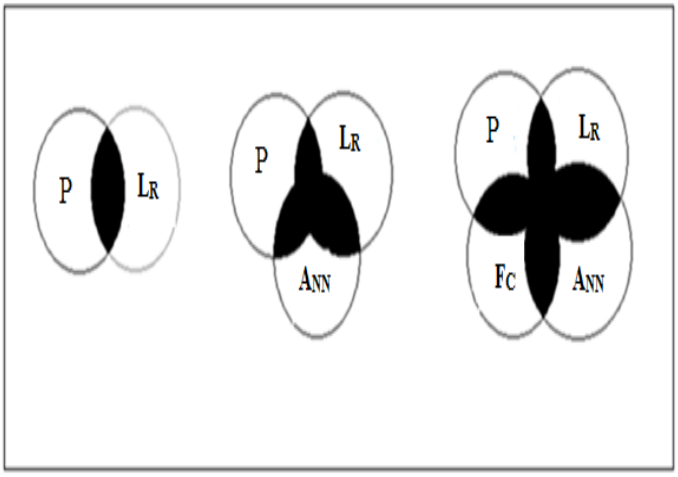
(C) Representation of Sets and its operation:-

Union Representation:-
Set P = {P_0, P_1, P_2, P_3}
Set L_{RI} = {L_{R0}, L_{R1}, L_{R2}, L_{R3}}
Set (P U L_R) = {P_0, P_1, P_2, P_3 U L_{R0}, L_{R1}, L_{R2}, L_{R3}}

Set A_{NN} = {A_{NN0}, A_{NN1}, A_{NN2}, A_{NN3}}
Set (P U L_R U A_{NN}) = {P_0, P_1, P_2, P_3, L_{R0}, L_{R1}, L_{R2}, L_{R3} U A_{NN0}, A_{NN1}, A_{NN2}, A_{NN3}}

Set F_C = {F_{C0}, F_{C1}, F_{C2}, F_{C3}}
Set (P U R_{OI} U C_{NN} U D_T) = {P_0, P_1, P_2, P_3, L_{R0}, L_{R1}, L_{R2}, L_{R3}, A_{NN0}, A_{NN1}, A_{NN2}, A_{NN3} U F_{C0}, F_{C1}, F_{C2}, F_{C3}}

Venn diagram



V CONCLUSION

The paradigm of crop yield prediction is one of the most essential needs of the hour. The increase in droughts and other crop failures have resulted in major loss of life and livelihood across the globe. The increase in the incidents of global warming and changing weather has led to an increased uncertainty over the crop yield. The crop yield has also been highly problematic to detect as the uncertainty looms large. There are certain approaches that have been effectively providing a prediction for crop yield but most of the approaches have been ineffective or have problems in accuracy of the prediction. This research has analyzed and effectively identified the various researches for the purpose of achieving an effective and accurate approach towards crop yield production. Therefore, an effective technique for predicting the future yield of crops is being formulated using Artificial Neural Networks and Fuzzy Classification. The devised idea will be elaborated further in our future editions.

VI ACKNOWLEDGEMENTS

It is our Great Pleasure to express our special thanks of gratitude to all our Authors and Professors for their contribution in our project for analysis and reporting of our work, also we would like to thank them who supported for important intellectual content also for article drafting and revising it.

VII 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] HosseinAghighi 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.

[7] MicheleMeroni 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] ZhihongZhuo et al," Regional Grain Yield Response to Climate Change in China: A Statistic Modeling Approach ", IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, VOL. 7, NO. 11, NOVEMBER 2014.

[9] RobertoLuciani, "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] DhivyaElavarasan et a, Sandhya," Crop Yield Prediction Using Deep Reinforcement Learning lModel for Sustainable Agrarian Applications ", IEEE Access, 2020.

****