CROP YIELD PREDICTION USING MACHINE LEARNING

Supriya Jagtap, Aakanksha Gaikwad, Vinay Dongare, Nilesh Bangar,Prof.Amar Chadchankar

Dept. Of Information Technology,

PCET's Nutan Maharashtra Institute of Engineering and Technology,Pune, India

ABSTRACT

To develop effective agricultural and food policies at the regional and global scales accurate predictions of crop yield are critical. Considering the present system including manual counting, climate smart pest management and satellite imagery, the result obtained aren't really accurate. This paper focuses mainly on predicting the yield of the crop by applying various machine learning techniques which plays a significant role as it has a decision support tool for Crop Yield Prediction (CYP). Crop yield prediction is an essential task for the decision-makers at national and regional levels (e.g., the EU level) for rapid decision-making. An accurate crop yield prediction model can help farmers to decide on what to grow and when to grow. The present researchdeals with a systematic review that extract and synthesize the features used for CYP and furthermore, there are avariety of methods that were developed to analyze crop yield prediction using artificial intelligence techniques. This paper explores various ML techniques utilized in the field of crop yield estimation and provided a detailed analysis in terms of accuracy using the techniques. According to our analysis, the most used features are temperature, rainfall, and soil type. Random Forest is the most widely used machine learning algorithm in thesestudies, and the other widely used algorithms are Long-Short Term Memory (LSTM) and K-Nearest Neighbor algorithm (KNN).

Keywords- *Agriculture, Random Forest, Crop yield prediction, are Long-Short Term Memory (LSTM), Machinelearning method.*

INTRODUCTION:

Machine learning, which is a branch of Artificial Intelligence (AI) focusing on learning, is a practical approach that can provide better yield prediction based on several features. Machine learning (ML) can determine patternsand correlations and discover knowledge from datasets. The models need to be trained using datasets, where the outcomes are represented based on past experience. The predictive model is built using several features, and as such, parameters of the models are determined using historical data during the training phase. For the testing phase, part of the historical data that has not been used for training is used for performance evaluation purpose.

Agriculture is the backbone of India's economy since its plays a vital role in the survival of every human and animal in India. The worldwide population is 7.9 billion and is predicted to increase to 8.5 billion by 2030, leadingto an extreme increase in demand for agricultural products. In the future, agricultural products will have higher demand among the human population, which will require efficient development of farmlands and growth in the yield of crops. Meanwhile, due to global warming, the crops were frequently spoiled by harmful climatic situations. A single crop failure due to lack of soil fertility, climatic variation, floods, lack of soil fertility, lack ofgroundwater and other

such factors destroy the crops which in turn affects the farmers. In other nations, the society advises farmers to increase the production of specific crops according to the locality of the area and environmental factors. The population has been increasing at a significantly higher rate, so the estimation and monitoring of crop production is necessary. Accordingly, an appropriate method needs to be designed by considering the affecting features for the better

The core objective of crop yield estimation is to achieve higher agricultural crop production and many established models are exploited to increase the yield of crop production. Nowadays, ML is being used worldwide due to its efficiency in various sectors such as forecasting, fault detection, pattern recognition, etc. The ML algorithms also help to improve the crop yield production rate when there is a loss in unfavorable conditions. The ML algorithms are applied for the crop selection method to reduce the losses of crop yield production irrespective of distracting environment.

METHODOLOGY

1. Data Pre-Processing

Data Preprocessing is a method that is used to convert the raw data into a clean data set. The data are gathered from different sources, it is collected in raw format which is not feasible for the analysis. By applying different techniques like replacing missing values and null values, we can transform data into an understandable format. The final step on data preprocessing is the splitting of training and testing data. The data usually tend to be split unequally because training the model usually requires as much data- points as possible. The training dataset is the initial dataset used to train ML algorithms to learn and produce right predictions (Here 80% of dataset is taken as training dataset).

2. Factors affecting Crop Yield and Production

There are a lot of factors that affects the yield of any crop and its production. These are basically the features that help in predicting the production of any crop over the year. In this paper we include factors like Temperature, Rainfall, Area, Humidity and Windspeed (Fig.1 shows the features).

3. Comparison and Selection of Machine Learning Algorithm

Before deciding on an algorithm to use, first we need to evaluate and compare, then choose the best one that fits this specific dataset. Machine Learning is the best technique which gives a better practical solution to crop yield problem. There are a lot of machine learning algorithms used for predicting the crop yield. In this paper we include the following machine learning algorithms for selection and accuracy comparison (TABLE1 shows the algorithm and its accuracy):

3.1) LSTM:- Long-short term memory (LSTM) is a specific recurrent neural network (RNN) architecture that is well suited to learn from experience to classify, process and predict time series with time lags of unknown size. LSTM neural networks are composed by units called memory blocks, and each memory block contains memory cells with self-connections storing the temporal state of the network, in addition to multiplicative units (gates) to control the flow of information. LSTM are useful when dealing with data with a temporal relationship and can learn to recognize temporally extended patterns in noisy sequences.

To train a LSTM the dataset is divided into a static set and a dynamic set. The static set is integrated

by the soil data, which for the time scales consider in this work soil properties do not change over time. The dynamic set is integrated by the monthly weather set, containing meteorological data with seasonal variability.

3.2) KNN:- A K-Nearest Neighbor algorithm is a data categorization approach that examines neighboring data points to determine which group or class a data point belongs to. The K-Nearest Neighbor algorithm examines a point on the grid at identifies which group or class it belongs to. To do so, it examines other data points in the vicinity and determines their class. The majority of the neighbors around the data point are allocated to the same class as the data point. K-Nearest Neighbor algorithm does not construct a data model. When we ask it to classify a data point, it calculates everything during the runtime. It is also known as 'Lazy Learner' because of this.

3.3) Random Forest:- Random Forest has the ability to analyze crop growth related to the current climatic conditions and biophysical change. Random forest algorithm creates decision trees on different data samples and then predict the data from each subset and then by voting gives better solution for the system. Random Forest uses the bagging method to train the data which increases the accuracy of the result. For our data,RF provides an accuracy of 92.81%. It is clear that among all the three algorithms, Random forest gives the better accuracy as compared to other algorithms.

4. Random Forest Model for Crop Prediction

Random forests are the aggregation of tree predictors in such a way that each tree depends on the values of a random subset sampled independently and with the same distribution for all trees in the forest. Random Forest used the bagging method to trained the data which increases the accuracy of the result. For getting high accuracywe used the Random Forest algorithm which gives accuracy which predicate by model and actual outcome of predication in the dataset. The predicted accuracy of the model is analyzed 91.34%.

5. Proposed System

Our proposed system is a web application which predicts name of the crop as well as calculate its corresponding yield. Name of the crop is determined by several features like temperature, humidity, wind-speed,rainfall etc. and yield is determined by the area and production. In this paper, Random Forest classifier is used forprediction. It will attain the crop prediction with best accurate values. (Fig.2 shows the System Architecture).

FIGURES AND TABLES

Table 1) Algorithm and its Accuracy

Algorithm	Accuracy
Random Forest	92.81407991690006
LSTM	90.56744638764848
KNN	87.46546543787506

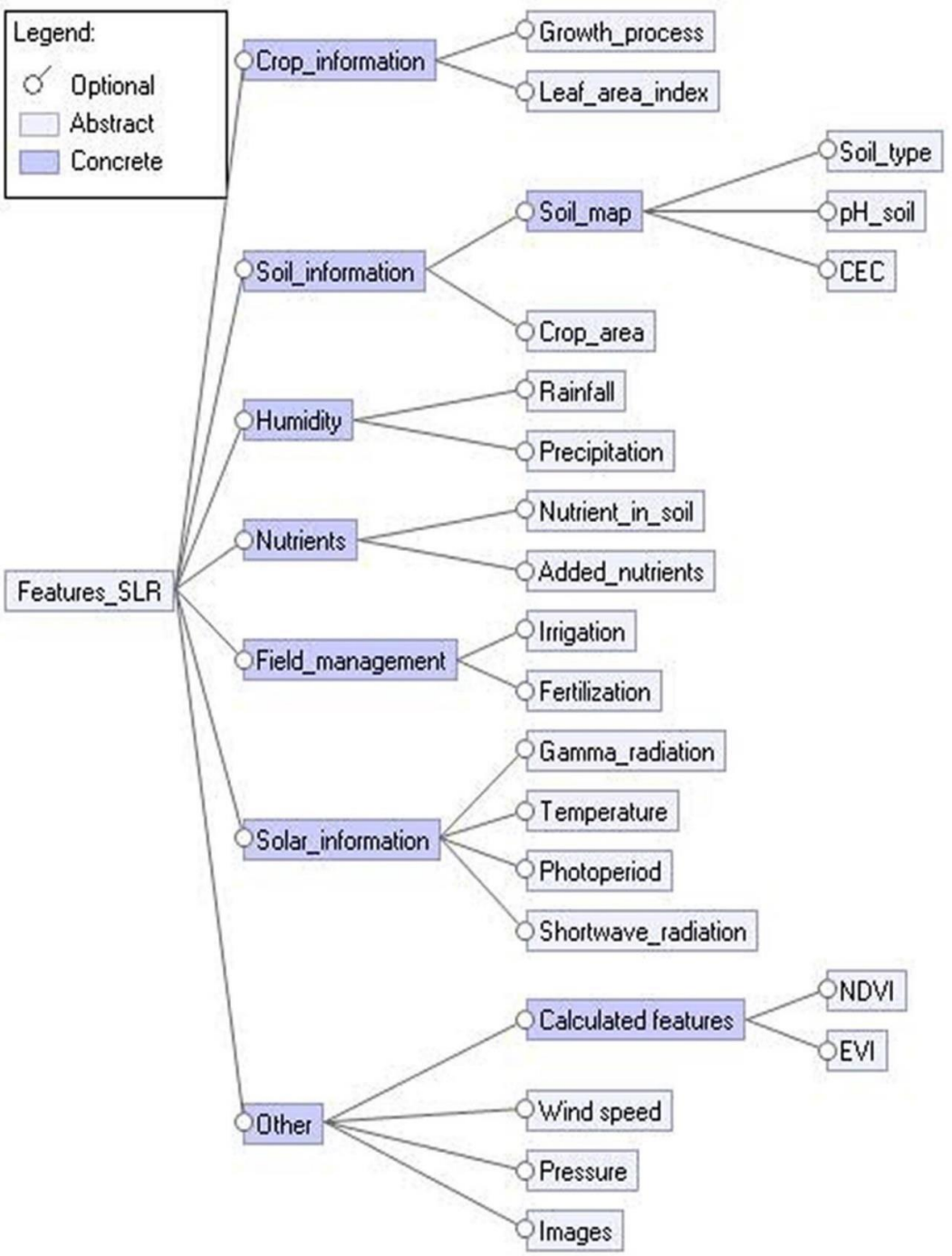

Fig. 1)Features used in dataset

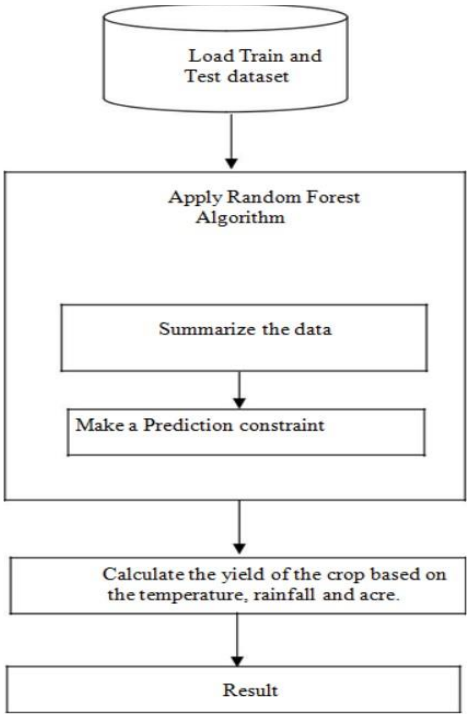

Fig 2) System Architecture

CONCLUSION:

The present research work discussed about the variety of features that are mainly dependent on the data availability and each of the research will investigated CYP using ML algorithms that differed from the features. The featureswere chosen based upon the geological position, scale, and crop features and these choices were mainly dependentupon the data-set availability, but the more features usage was not always giving better results. Therefore, findingthe fewer best performing features were tested that also have been utilized for the studies. Studies stated that models with more features did not always provide the best performance for the yield prediction. To find the best performing model, models with more and fewer features should be tested. Many algorithms have been used in different studies. The results show that no specific conclusion can be drawn as to what the best model is, but theyclearly show that some machine learning models are used more than the others. The most used models are the random forest, K-Nearest Neighbors Algorithm, Long-short term memory (LSTM). Most of the studies used a variety of machine learning models to test which model had the best prediction. Ultimately, the experimental study showed the combination of ML with the agricultural domain field for improving the advancement in crop prediction. Additionally, in the crop yield estimation, fertilizer should also be considered for executing soil forecasts that agriculturalist to make a better judgment based on the situation of low crop yield estimation. Basedon the outcomes obtained for the study further we need to build and develop a model based on DL for CYP.

REFERENCES:

[1] P.Priya, U.Muthaiah M.Balamurugan.Predicting yield of the crop using machine learning algorithm.International Journal of Engineering Science Research Technology.

[2] J.Jeong, J.Resop, N.Mueller and team. Random forests for global and regional crop yield prediction. PLoS ONEJournal.

[3] Narayanan Balkrishnan and Dr. Govindarajan Muthukumarasamy. Crop production Ensemble MachineLearning model for prediction. International Journal of Computer Science and Software Engineering (IJCSSE).

[4] S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh. Machine learning approach for forecasting crop yield basedon climatic parameters. International Conference on Computer Communication and Informatics (ICCCI).

[5] Shweta K Shahane , Prajakta V Tawale. Prediction On Crop Cultivation. International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE) Volume 5, Issue 10, October 2016.

[6] D Ramesh ,B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology. [7]Subhadra Mishra, Debahuti Mishra, Gour Hari Santra. Applications of Machine Learning Techniques in Agricultural Crop Production. Indian Journal of Science and Technology, Vol 9(38), DOI:10.17485/ijst/2016/v9i38/95032, October 2016.

[7] Aruvansh Nigam, Saksham Garg, Archit Agrawal “Crop Yield Prediction using ML Algorithms “,2019

[8] Leo Brieman, “Random Forests”, 2001

[9] Priya, P., Muthaiah, U., Balamurugan, M.”Predicting Yield of the Crop Using Machine Learning Algorithm”,2015

[10] Mishra, S., Mishra, D., Santra, G. H., “Applications of machine learning techniques in agricultural crop production”,2016