International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Review on Crop Yield Predication Using Machine Learning Techniques

*Sheela Asare Chinchmalatpure[a], Dr. Phani KumarS.[b]**

[a] Reserch Scholar , Gitam University Hyderabad , India
[b] Professor of Computer Department , Gitam University Hyderabad, India

A B S T R A C T

Agriculture is one of India's most important and least well-paid occupations. Artificial intelligence approaches have been used to study crop yield prediction in a variety of ways. Machine learning is a useful decision-making tool for predicting crop yields, as well as for determining which crops to plant and what to do during the crop's growth season. The current state of machine learning and its use in agriculture will be the topic of our discussion. In order to extract and synthesize the approaches and attributes used in agricultural production prediction research, a systematic review was conducted in this work.

Keywords: Agriculture, Neural Network, Support Vector Machine, Crop yield prediction, Machine learning method.

1. Introduction

Agricultural is the backbone of the economic system of a India, not only providing food and raw material, but providing employment opportunities for a large percentage of the population. There is an increase in world population, which is affecting land, water and resources, reducing the number of resources we can use to continue the supply-demand network. So, we need a smarter approach and become more efficient about how we farm to be most productive. Artificial techniques are used in the agricultural sector to increase the accuracy as well as to find solutions to difficult problems. Machine learning (ML) is one of the fastest growing areas in agriculture.ML can be characterized as the logical technique that will permit machines the capacity to learn without programming the gadgets. An ML model might be descriptive or predictive, depending on the study problem and research aims. Predictive models are used to generate future predictions, whereas descriptive models are used to extract information from collected data and describe what has happened. [16] (Alpaydin, 2010).

Nowadays, machine learning is being utilized worldwide due to its effectiveness in several sectors. Farmers frequently examines the crops both before and after harvest for the existence of special features, the proportion of damage, and other aspects that would affect the end viable product and retail price using a machine learning algorithm.

All crop varieties are categorized primarily by soil type and season. In our country, the monsoon plays a significant part in determining the role of agriculture this year. It was feasible to recommend crops to farmers based on their soil type and climate. Crop output is influenced by four key factors: soil fertility, water availability, climate, and diseases or pests. As a result of globalization, climatic conditions are currently unpredictable. As a result, farmers are having difficulty forecasting weather and crop conditions using meteorological data. Crop predation is accomplished out using a machine learning technique that combines historical and current data from a specific month. To aid the country's agricultural progress, several researchers have employed machine learning methodologies such as SVM, linear regression, KNN, decision tree, neural network, random forest, and gradient boosting tree.

We conducted a comprehensive literature study to acquire an overview of what has been done on the application of machine learning in crop yield prediction. Rest of the paper is organized as follows. Section 2 gives the block diagram and methodology applied for predicting crop yield. Section 3 focuses on related work and Section 4 concludes the paper with comparative analysis.

1.1. Block Diagram

Crop yield prediction is an important agricultural challenge. For agricultural risk management and future predictions, accurate knowledge regarding crop yield history is important. The following are the steps involved in crop yield prediction using machine learning.

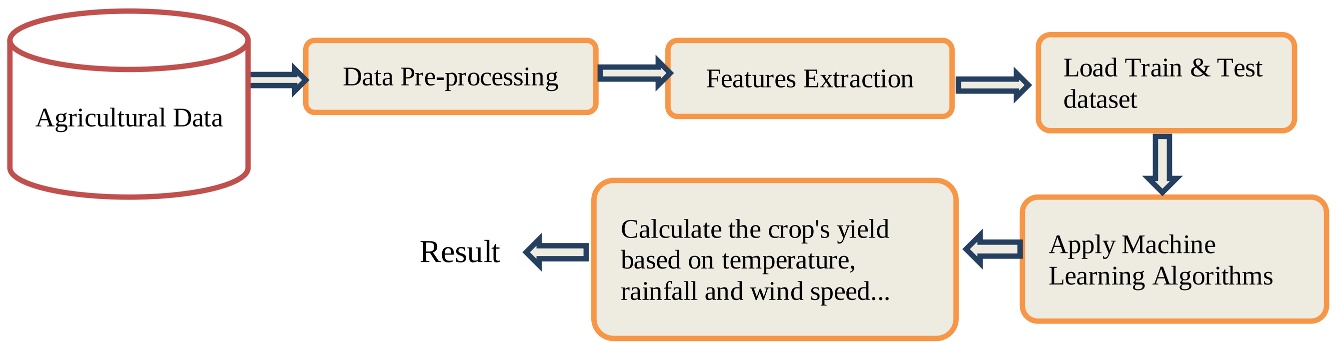

Fig 1:- Block diagram of the crop yield prediction using ML algorithms

Step 1:- Load the agricultural dataset with a set of parameters.
Step 2:- The data is pre-processed to remove any noisy data.
Step 3:- The pre-processed data is subjected to a feature extraction procedure that extracts information such as soil information, humidity, wind speed and nutrients among other things.
Step 4:- The data is separated into training and testing sets in order to prepare the dataset.
Step 5:- Machine learning algorithms are used to forecast the best crop to plant.
Step 6:- The test dataset evaluates the model's performance.

Machine learning approaches have the potential to handle large non-linear problems autonomously using datasets from a variety of sources, with or without human intervention, allowing for better decision-making and informed actions in real-world situations. The most applied ML algorithms are SVM, linear Regression, random forest and Neural Network.

Support Vector Machine Regression:-

SVR was first introduced by Drucker et al. as a supervised learning technique based on Vapnik's support vectors notion [1]. SVR attempts to reduce error by locating the hyperplane and decreasing the range between expected and observed values. Using equation (1) to reduce the value of w

$$\min \|w\|^2 + C\sum_i^n(\xi_i^+ + \xi_i^+) \quad (1)$$

The empirical error is represented by the summation portion. As a result, the following equation (2) is used to reduce the error.

$$f(x) = \sum (\alpha_i^* + a_i)\, k\,(x, x_i) + B \quad (2)$$

Where (α_i^*, a_i) represents Lagrange multiplier $(\alpha_i^*, a_i) \geq 0$ $\boldsymbol{K}$ represents the kernel function and $\boldsymbol{B}$ represents the bias term. SVR is more accurate in predicting performance and it is effective at resolving high-dimensional feature regression problems [2].

Linear Regression:

Linear regression is one of the most fundamental types of machine learning, in which we train a model to predict the behaviour of your data based on a set of variables. As the name implies, linear regression requires the two variables on the x- and y-axes to be linearly connected. The predictor variable X is used to predict a quantitative response Y using linear regression. A linear regression equation can be written mathematically as:

$$y = a + bx$$

Where a and b are determined by the following formulas:

$$b(slope) = \frac{n\sum xy - (\sum x)(\sum y)}{n\sum x^2 - (\sum x)^2} \quad a(intercept) = \frac{n\sum y - b(\sum x)}{n}$$

Here, x and y are two variables on the regression line.
b = Slope of the line
a = y-intercept of the line.
x = Independent variable from dataset
y = Dependent variable from dataset

We get the best fit line once we identify the $a(intercept)$ and $b(slope)$ values. So, when we use our model to make a forecast, it will predict the value of y given the value of x.

Random Forest:

A random forest (RF) is a machine learning technique for solving classification and regression problems. There are several decision trees in a random forest algorithm. Bagging or bootstrap aggregation are used to train the 'forest' created by the random forest method. Bagging is a machine learning approach that uses an ensemble meta-algorithm to increase accuracy. The RF algorithm determines the outcome based on the decision trees' predictions. It makes predictions by averaging or averaging the output of various trees. The precision of the outcome improves as the number of trees grows.

2. Related Work

Liakos et al. 2018 [3] published a review paper on applications of machine learning in agricultural production systems. The analysis was conducted on crop management, disease detection, weed detection, crop quality, species recognition, livestock management, water management and soil management. The purpose of this paper is to show how machine learning technology can benefit agriculture.

Gandhi and Armstrong 2016 [4] published a review paper on the use of data mining in the agriculture industry, with a focus on decision-making .The application of ANN, Bayesian networks, and SVM to explore the links between various climatic and other factors on agricultural production is discussed in this paper. They came to the conclusion that more research is needed to determine how data mining may be implemented into complex agricultural databases.

Reshma et al 2020 [5] proposed a method for soil content analysis and crop yield prediction based on IoT-based classification techniques. Data mining techniques are used to classify soil into low, medium, and high categories in order to forecast crop output using available datasets. The data is analyzed using a decision tree and a support vector machine.

Sun et al. 2019 [6] used the Deep CNN-LSTM approach to estimate soybean yields. According to the most recent research in this area, CNN could investigate more spatial information and LSTM could uncover phonological features, both of which are important in crop production prediction. All of the training data was combined and turned into histogram-based tensors for deep learning using the Google Earth Engine (GEE). The results of the experiments show that the proposed CNN-LSTM model can outperform over the pure CNN or LSTM model.

S Dhakshina Kumar et al. 2020 [7] proposed a microcontroller-based machine vision technique for tomato grading and sorting using SVM technology. The tomato classifications were accomplished in three stages utilizing digital pictures of samples gathered in an experimental setting and a microcontroller. As it comprises an unambiguous perception of the faults, the crop data was categorized based on the texture, shape, and color of patterns on the diseased surface.

Leonardo et al. 2021[8] published a study that used post-processing data and machine learning approaches to forecast sugarcane yield using CAN (controller area network) data from sugarcane harvesters. Fuel consumption, engine rotation, engine power, and specific fuel consumption (SFC) were employed as data input parameters in training and testing random forest (RF), multiple linear regression (MLR), and artificial neural network (ANN) predictive models using CAN data. The RF model with a mean absolute percent error (MAPE) and root mean square error (RMSE) produced the best results for estimating sugarcane output based on engine parameters.

Vaishali et al. 2020 [9] published a paper that used several machine learning methods to predict Mustard Crop yield from soil data in advance. Multinomial logistic regression, K-nearest neighbour (KNN), ANN, random forest, and Naive Bayes were used to forecast mustard crop output in advance from soil investigation. The most reliable strategies for predicting mustard crop yields were determined to be KNN and ANN in this study. Farmers will be able to anticipate yield in advance using these effective machine learning approaches based on soil data.

Bhanu et al 2020 [10] proposed a study on an IoT-based intelligent system for agriculture and explores the application of machine learning in a cloud platform. This work examines the current state of IoT in agriculture through key literary works, as well as developments in IoT, cloud platforms, agricultural applications, IoT-applications, and recent difficulties.

3. Conclusion:

The findings revealed that, depending on the scope of the research and the availability of data, the selected papers employ a number of features. The features chosen are determined by the dataset's availability and the research's goal. Additionally, studies have shown that models with more characteristics may not always deliver the highest yield prediction performance. Models with more and fewer features should be evaluated to determine the best performing one. Several algorithms have been employed in various research. The findings reveal that no definitive conclusion can be taken about which model is the best, but they do show that some machine learning models are utilized more frequently than others. . Random forests, neural networks and linear regression are the most commonly employed models.

4. Comparative table:

Paper Title	Authors	Methodology	Features	Performance Metric
Smart Farming System: Crop Yield Prediction Using Regression Techniques [11]	Ayush Shah, AkashDubey, VisheshHemnani, Divye Gala , D. R. Kalbande	SVM, Multivariate Polynomial Regression, Random Forest	temperature and rainfall	SVM R-squared =0.968. RMSE = 5.48 MAE = 3.57 MdAE = 1.58
Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil [12]	Raí A. Schwalbert, Telmo Amado , GeomarCorassa , Luan Pierre Pott , P.V.Vara Prasad, Ignacio A. Ciampitti	Multivariate OLS linear regression, Random forest and LSTM neural networks	satellite imagery and weather data	LSTM (DOY 16) MAE= 0.52 RMSE= 0.68 MSE=0.46
Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches [13]	Xinlei Wang , Jianxi Huang , Quanlong Feng and Dongqin Yin	LSTM & CNN	meteorological ,remote sensing and soil features.	R= 0.77 RMSE =721 kg/ha
Crop Yield Predication using Deep Neural Networks [14]	Saeed Khaki, Lizhi Wang	Deep Neural Network	Soil and weather data	RMSE=10.55 (Training) RMSE=12.79 (validation)
Sugarcane Yield Grade Predication using Random Forest and Gradient Boosting Tree techniques [15	Phusanisa Charoen-Ung, PraditMittrapiyanuruk	Random Forest classification, and Gradient Boosting tree classification	soil type, plot area, groove width, sugarcane class and type	Accuracy RF= 71.83% GBT =d 71.64%

References:

1) H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, ''Support vector regression machines,'' in Proc. Adv. Neural Inf. Process. Syst., 1997, pp. 155–161
2) Singh, Abhilash, VaibhavKotiyal, Sandeep Sharma, Jaiprakash Nagar, and Cheng-Chi Lee. "A Machine Learning Approach to Predict the Average Localization Error With Applications to Wireless Sensor Networks." *IEEE Access* 8 (2020): 208253–208263.
3) Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: a review. Sensors (Switzerland) 18 (8). https://doi.org/10.3390/ s18082674.
4) Gandhi, N., Armstrong, L.J., 2016b. "A review of the application of data mining techniques for decision making in agriculture". In Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, https://doi.org/10.1109/ IC3I.2016.7917925.
5) R. Reshma, V. Sathiyavathi, T. Sindhu and K. Selvakumar 2020 "IoT based Classification Techniques for Soil Content Analysis and Crop Yield Prediction ". In Proceedings of the Fourth International Conference on I-SMAC , IEEE 2020 ,pp 157-160
6) J. Sun, L, Di, Z. Sun, Y. Shen, and Z. Lai, "County-level soybean yield prediction using deep CNN-LSTM model," Sensors, vol. 19, no. 20, pp. 4363, 2019.
7) S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, pp.103090, 2020

8) Leonardo Felipe Maldaner , Lucas de Paula Corrˆedo , Tatiana Fernanda Canata, "Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches". https://doi.org/10.1016/j.compag.2020.105945
9) V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, "Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis," Journal of Scientific Research, vol. 64, no. 2, 2020.
10) Bhanu K N, Jasmine H J, Mahadevaswamy H S, "Machine learning Implementation in IoT based Intelligent System for Agriculture" 2020 International Conference for Emerging Technology (INCET) Belgaum, India. IEEE-2020.
11) Ayush Shah, AkashDubey, VisheshHemnani, Divye Gala and D. R. Kalbande," Smart Farming System: Crop Yield Prediction Using Regression Techniques", Proceedings of International Conference on Wireless Communication, , Springer 2018
12) Raí A. Schwalbert, Telmo Amado , GeomarCorassa , Luan Pierre Pott , P.V.Vara Prasad, Ignacio A. Ciampitti," Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil", https://doi.org/10.1016/j.agrformet.2019.107886 2019 Elsevier
13) Xinlei Wang , Jianxi Huang , Quanlong Feng and Dongqin Yin ," Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches", http://www.mdpi.com/journal/remotesensing 2020.
14) Saeed Khaki, LizhiWang ,"Crop Yield Prediction Using Deep Neural Networks", " Frontiers in plant science, vol. 10, pp. 621, 2019. https://doi.org/10.3389/fpls.2019.00621
15) Phusanisa Charoen-Ung, PraditMittrapiyanuruk ," Sugarcane Yield Grade Prediction using Random Forest and Gradient Boosting Tree Techniques", 15th International Joint Conference on Computer Science and Software Engineering (JCSSE), IEEE 2018
16) Alpaydin, E., 2010. Introduction to Machine Learning, 2nd ed. Retrieved from https:// books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq= introduction+to+machine+learning&ots=T5ejQG_7pZ&sig=0xC_ H0agN7mPhYW7oQsWiMVwRnQ#v=onepage&q=introduction to machine learning&f=false
17) C.G. Anupama, C. Lakshmi ," A comprehensive review on the crop prediction algorithms" https://doi.org/10.1016/j.matpr.2021.01.549 2021 Elsevier.
18) Janmejay Pant , R.P. Pant , Manoj Kumar Singh , DeveshPratap Singh , Himanshu Pant ," Analysis of agricultural crop yield prediction using statistical techniques of machine learning", 2021 Elsevier.