Crop Yield Prediction Using Machine Learning Algorithms

Mr. Manjunath K N
Dept. of CSE, RVITM
Bengaluru, India

Ms. Pallotu Devisri
Dept. of CSE, RVITM
Bengaluru, India

Mr. Bysani Balaji
Dept. of CSE, RVITM
Bengaluru, India

Ms. Shravya Suresh
Dept. of CSE, RVITM
Bengaluru, India

Ms. Sameeksha M
Dept. of CSE, RVITM
Bengaluru, India

Abstract— **Agribusiness is the primary source of income for the majority of India's population. Climate, weather, and other natural factors are currently seriously jeopardizing the long-term viability of agriculture. Since it has a decision help instrument for Crop Yield Prediction (CYP), which remembers direction for which yields to create and what to do all through the collect improvement season, machine learning (ML) assumes a significant part. A precise survey that focuses on and combines the characteristics used for CYP is the focus of the ongoing review. The current review's main focus is on a variety of computerized reasoning techniques developed to examine rural produce anticipation. The brain organization's fundamental obstructions are diminished horticultural result check adequacy and lower rate bumble. Managed learning computations were unable to account for the intricate relationship between information and outcome variables when evaluating or organizing natural goods. A lot of research was suggested for agricultural events, such as crop order in relation to growth stage and temperature, crop illness clustering, and crop output indicators, in order to develop a precise and effective model for crop classification. The accuracy of a number of machine learning (ML) algorithms used to predict agricultural output is the subject of this study.**

Keywords—*Machine learning algorithms, Crop Yield Prediction (CYP).*

I. INTRODUCTION

Agriculture serves as the foundation of the Indian economy because it is necessary for both humans and animals to survive [1]. The global population will rise up to 4.9 billion in 2030. As a result, farm products will see significant increases in consumption. Future demand for agricultural products will call for more crop production and efficient land expansion. However, severe weather frequently destroyed harvests as a result of global warming [2].When a single crop fails for several reasons, including less adequacy of soil maturity, environmental variation, storms, a lack of freshwater, and others, farmers suffer. Based on location and climate, some societies encourage farmers to enhance the production of particular goods [3]. It is essential to measure and monitor agricultural results due to the rapid population growth [4]. As a consequence of this, the influencing factors ought to be taken into consideration when developing a competent method for selecting additional evolved crops while taking into account unusual vacillations [5].

Increasing agricultural output is the primary goal of agricultural yield evaluation, which is accomplished by a variety of deeply rooted processes. Due to its usefulness in a wide range of applications such as gauging, malformation detection, design recognition, and so on, machine learning (ML) is widespread worldwide. The crop output generation rate is also helped by the ML formulas in the event of a mishap brought on by unfavourable circumstances. The harvest option method employs ML computations to reduce miscalculations in agricultural output generation despite the presence of distracting external factors. The continuing model used SVM to organize the data for a particular crops based on the surface, shape, and tone of instances on the weakened surface [6] because it has a good sense of the flaws.

*A.*CNN was used in a recent method to reduce overall error and improve agricultural output prediction [7]. Similar to this, the current method, which combined a time series model with Back Propagation Neural Networks (BPNNs) and a smaller dataset size, was unsuccessful because fewer instances were used for prediction [8, 9].

B. When selecting power and accuracy, ML techniques were utilized. Using ML, it is possible to address the information output connection in agricultural prediction in a number of useful ways. For crop selection, weather conditions assessment, yield forecasting, amazing water system frameworks, crop infection forecasting, determining the basic assistance cost, and other gardening endeavors, various machine methods are utilized. These strategies will result in a decrease in farmers' input efforts and an increase in farming output. In addition, technological and computer advancements were precise due to their significant role in data utilization [10].

Fig. 1 Agriculture Harvesting

II. LITERATURE REVIEW

1. Predicting yield of the crop using machine learning algorithm.

The country's economic success is significantly influenced by agriculture. A significant threat to agribusiness has emerged as a result of weather and climate change. A crucial method for coming up with solutions that are both useful and effective to this problem is machine learning, or ML. Forecasting crop output using historical data like crop yield, soil factors, and weather conditions is known as crop output prediction. In this study, projections of agricultural output are predicted using the Random Forest method based on available data. The examples were used to test the models, which were developed with real-world Tamil data. Before growing on arable land, the farmer will use the prediction to anticipate production. Random Forest, a complex and broadly utilized directed ML technique, is utilized to foresee future cultivating yield accurately.

2. Applications of machine learning techniques in agricultural crop production: a review.

The purpose of this report was to look into the study's findings about how ML techniques can be used in agriculture to grow crops. Methods for gathering and analysing data: This is a novel approach to crop output control in farming. For crucial strategy decisions like import-trade, cost, marketing transportation, and so on, the Directorate of Financial elements and Insights provides precise and timely agricultural output forecasts. In any case, it is essential to keep in mind that these previous evaluations are inaccurate due to the need for extensive explanation in light of a variety of biased standards. Consequently, a logical and analytically sound estimate of agricultural output is required. A huge amount of data has been created as computers and data storage technology have developed. Findings: New approaches, such as machine learning, have been developed to integrate data expertise with farm output evaluation due to the difficulty of extracting complex information from unstructured data. This study looked at these novel strategies for dealing with large correlations between them and the various data collection factors. Application/Improvement: Among the methods are relapse examination, decision trees, artificial neural networks, information fuzzy networks, and Bayesian belief networks. The Markov chain model, k-means grouping, k nearest neighbour, and support vector machine were fully explained using agriculture.

3. A Model for Prediction of Crop Yield.

Another area of research in agricultural production analysis is information gathering. Production anticipation is a crucial issue in agribusiness. How much he could expect to collect would need to be known by each grower. Beforehand, creation assumptions depended on the farmer's past involvement in a specific yield and region. The fundamental issue of output prediction cannot be solved with the knowledge that is currently available. Techniques for extracting data are the best choice for this purpose. To forecast agricultural output for the following year, a variety of Data Mining methods are utilized and analysed in the agricultural sector. Provable data are used to develop and test a method for forecasting agricultural efficiency. To accomplish this, affiliation rule mining on farming information is utilized. The objective of this examination is to make an expectation model that can be utilized to foresee future cultivating results. This brief investigation of agricultural production forecasting employing an association-based information mining strategy focuses on the Indian state of Tamil Nadu. The results show that the proposed approach accurately predicts agricultural output.

4. Agricultural crop yield prediction using artificial neural network approach

By thinking about different parameters which are interlinked with the climate and that affects the temperature in different areas of the world. The production of food is directly impacted by the patterns of these weather conditions. The relationship between farming output and large-scale climatological events has been the subject of numerous studies. It has been demonstrated that artificial neural networks are effective forecasting and modelling tools for increasing their effectiveness. The best yield is determined by yield forecast using a number of land and weather factors. Take into consideration the soil's type, pH, nitrogen, phosphorus, potassium, natural carbon, manganese, copper, iron, density, weather, precipitation, and humidity, as well as calcium, magnesium, sulphur, and manganese. We used an artificial neural network as a result (ANN).

5. Predictive ability of machine learning methods for massive crop yield prediction.

For agricultural planning purposes, the accurate output calculation for the various commodities recalled is a significant issue. A crucial method for coming up with solutions that are both useful and effective to this problem is machine learning, or ML. For crop forecasting, a number of ML methods have been looked at to find the best one. In general, there aren't enough products and methods studied, and there aren't many data on farm planning. This study examines the accuracy of ML and linear regression algorithms in predicting food output from ten farming datasets. M5-Prime relapse trees, support vector regression, k-closest neighbour, and perceptron multi-facet neural networks were all evaluated. The models were approved by four exactness tests: the correlation factor (R), the standardized mean absolute error (MAE), and the root mean square error (RMS). The models were built with real-world data from a water system zone in Mexico. The models were evaluated using tests from two different years. The M5-Prime and k-closest neighbour techniques have the least run of the mill RMSE blunders (5.14 and 4.91), the most reduced average RRSE mistakes (79.46% and 79.78%), and the most elevated huge normal affiliation factors (0.41 and 0.42, individually). M5-Prime is an excellent tool for crop yield prediction in rural planning because it provides the most models with the fewest errors.

III. METHODOLOGY

For optimal prediction, the majority of current CYP models make use of KNN regression, random forests, and neural networks. A number of machine learning (ML) techniques were also utilized.

The current study on farm output forecasts using machine learning encounters the following issues:

1. Due to their complexity, ML systems were very expensive to develop, maintain, and fix. Even though the information and outcome information used in the ML method to predict wheat and mustard harvest production were included, it did not produce better results.
2. The relapse model was unable to accurately hypothesis complicated information, such as preposterous value and nonlinear information, due to the direct link between the borders.
3. For yield expectation and association, existing K-NN models were used, but KNN's unpredictable and highly adaptable issues hampered their presentation. They were used in a region model, which made it harder to organize things and stacked the information stream.

4. Due to a lack of evidence to predict agricultural production, no appropriate judgment was made during the classification process.

A.Advantages:

1. According to the study, CNN, LSTM, and DNN were the most frequently utilized formulas; But CYP needed to be improved.
2. The current study demonstrates a number of existing models that execute models for the most accurate prediction of agriculturaloutput while taking into account temperature and atmospheric variables.
3. In conclusion, the practical study demonstrated that harvest prediction progress was enhanced when ML and the farming area field were combined.

B.Disadvantages:

1. The output forecast and classification were performed using current K-NN models, but their performance was subpar due to KNN's nonlinear and extremely adaptable issues.
2. This study focuses primarily on the actual application and evaluation of machine learning methods. To produce a consistent pattern, the method described here examines the disjointed data from the temperature and precipitation records. Instead of focusing on each viewpoint individually, crop output forecasting examines the components as a whole.

IV. PROPOSED SYSTEM ARCHITECTURE

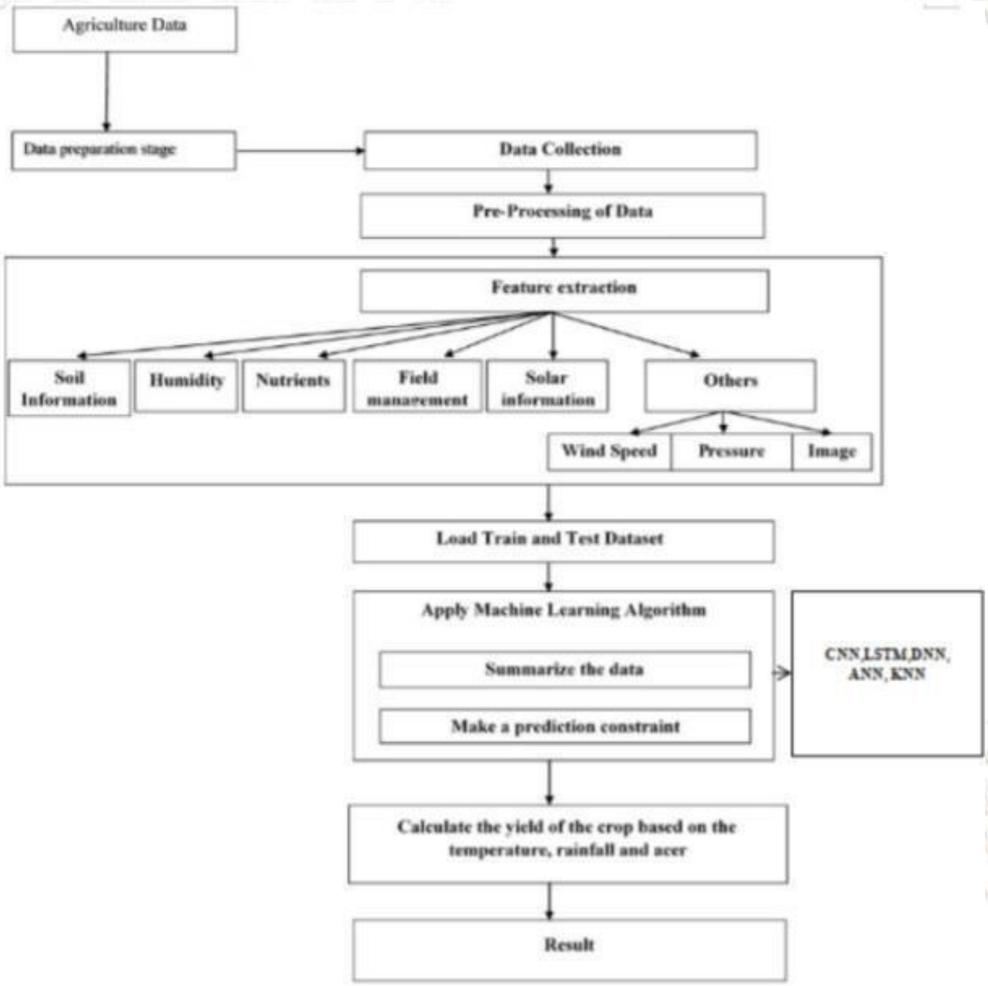

Fig. 2 Proposed Architecture

(i) Upload Crop Dataset

The information about agricultural output is used to predict the name and yield of the produce in the arrangement and recurrencecalculations.

(ii) Pre-process Dataset

The Random Forest Regressor was found to have the highest accuracy for output predictions using data from the Indian government. The sequence model, which is a Simple Recurrent Neural Network, outperforms the LSTM when it comes to rain forecasting. An estimate of the output for a particular location can be made by using temperature, moisture, and other statistics like season and area.

(iii) Train Machine Learning

Based on the commodities produced in the district, this focuses on district-specific production forecasts. Each agricultural region's output is determined by selecting the crop with the highest yield.

(iv) Upload Test Data &Predict Yield

At the point when all lines are consolidated, the outcomes show that Random Forest is the best indicator. In addition to assisting farmers in selecting the most suitable produce for the upcoming season, this will help to connect the technical and farming sectors.

V. IMPLEMENTATION ALGORITHMS

(i) Logistic Regression: Controlled calculating is used in the logistic regression order approach to estimate the likelihood of an objective variable. There are only two possible groups for the objective or ward variable due to its binary nature. The LR method is accurate 87.8% of the time when applied to our dataset.

(ii) Naive Bayes: As per the Naive Bayes classifier, the presence of different features doesn't correspond with the presence of one class component. The Naive Bayes model is simple to build and works best with extremely large data sets. Due to its simplicity, Naïve Bayes outperforms even the most extraordinary ranking techniques. It exactly equates to 91.50%.

(iii) Random Forest: Random Forest can be used to compare current weather and environmental change to crop development. The Random Forest method uses information from each subgroup to make predictions, creates option trees from different information tests, and then chooses the best answer for the framework. The destroying technique is utilized to orchestrate the information in Random Forest, and the outcome is more exact. For our data, RF provides an accuracy of 92.81 percent.

(iv) CNN: A deep learning network that learns directly from input is called a convolutional neural network, also known as a CNN or ConvNet. CNNs are extremely helpful in identifying picture patterns that can be used to distinguish between objects, groups, and divisions. Audio, time series, and signal data can all be categorized with their assistance.

(v) LSTM: Long short-term memory (LSTM) structures are utilized in Deep Learning. It is a kind of recurrent neural network (RNN) that can learn about situations over long periods of time and is useful for putting forecasting tasks into categories. LSTM can manage entire information groups rather than just single information centres like images because it has input connections.

(vi) DNN: Deep neural networks are a powerful class of machine learning computations because they build neural network levels along the depth and breadth of simpler designs.

(vii) KNN: The k-nearest neighbour's method (KNN or k-NN), a non-parametric supervised learning classification, uses nearness to characterize or predict the collection of a single piece of information. It can be used to solve characterisation or recurrence problems, but it is most commonly used as a clustering method because it relies on the possibility that themes that are similar can be seen as being in close proximity to one another.

VI. EXPERIMENTAL RESULTS

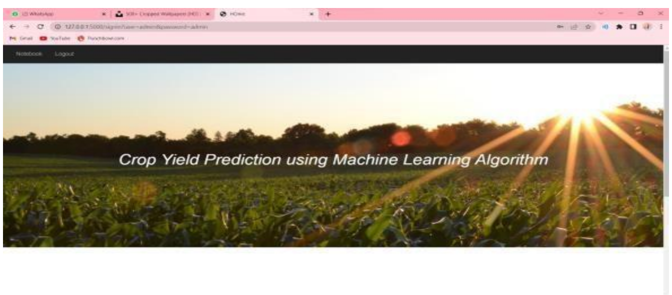

Fig. 3 Home Page front end GUI

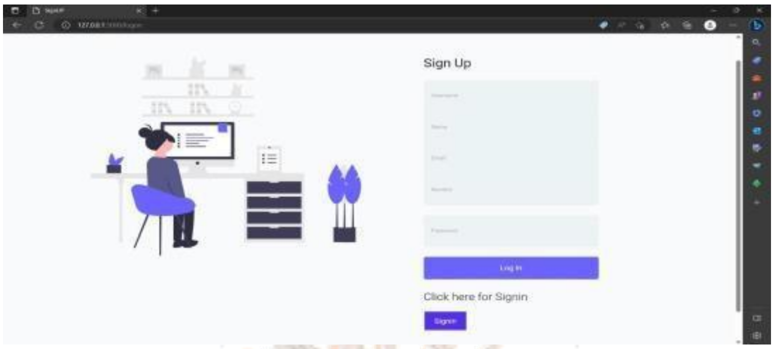

Fig. 4 Sign up Page

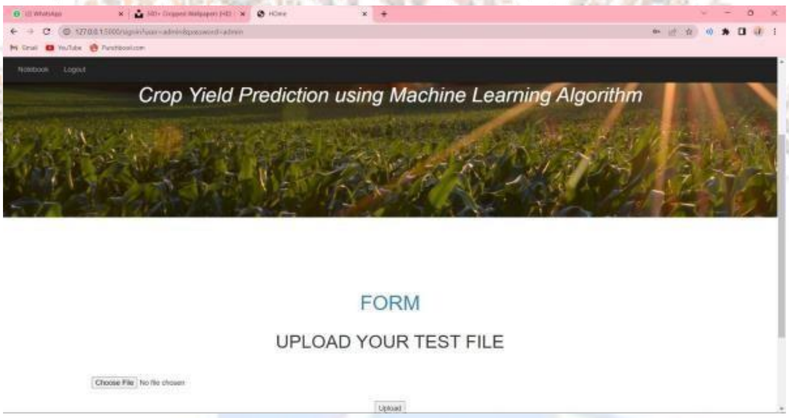

Fig. 5 Upload Image

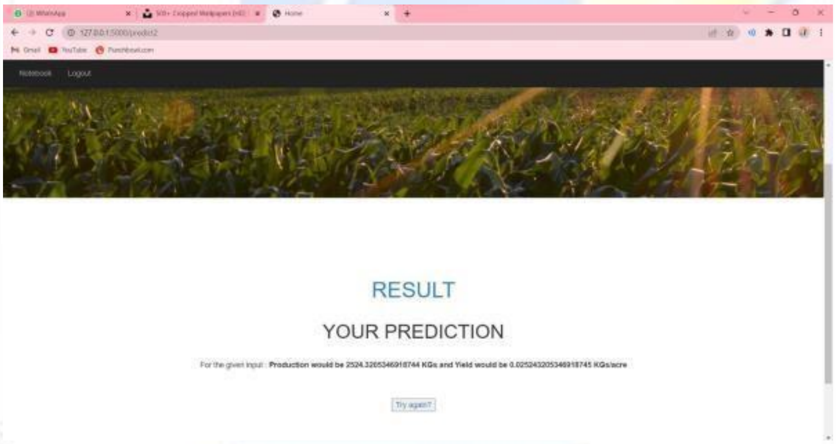

Fig. 6 Prediction result

VII. CONCLUSION

The ongoing evaluation addressed various aspects that are fundamentally dependent on information accessibility, and each study would examine CYP using distinct from the highlights ML methods. The selection of the highlights was influenced by the ease of data collection; However, better outcomes were not always achieved by using more components. The ground area, dimensions, and yield characteristics were used to select the components. Consequently, the characteristics with the lowest scores were investigated and utilized in the tests. Neural networks, random forests, and KNN recurrence techniques were utilized in the majority of current CYP models; However, the best prediction was also made using a variety of ML methods. CNN, LSTM, and DNN were found to be the most commonly used formulas by the investigation; nonetheless, CYP required extra development. The current study shows a variety of existing models that take into account weather, meteorological conditions, and predicting agricultural output. Lastly, the experimental investigation demonstrated that harvest prediction success could be improved by combining ML and the farming space field. However, it was anticipated that highlight selection would advance further in light of the effects of climate change on agriculture.

Additional express treatment was expected for the significant potential that ought to be the point of convergence of ensuing assessments, for instance, the underlying deferral to line geological areas. A random part of the model is created using ML, and then characteristics from deterministic crop models are used to get excellent quantifiable CO2 treatment. Additional exploration would make assessments of farming results in the event that the recently expressed objectives were followed. In addition, compost ought to be incorporated into the evaluation of crop yields in order to conduct soil numbers. This will enable agriculturalists to make better choices when crop yields are low. A DL-based model for CYP ought to be developed on the basis of the findings of the research.

VIII. REFERENCES

[1] R. Ghadge, J. Kulkarni, P. More, S. Nene, and R. L. Priya, "Prediction of crop yield using machine learning," Int. Res. J. Eng.Technolgy, vol. 5, 2018.

[2] F. H. Tseng, H. H.Cho, and H. T. Wu, "Applying big data for intelligent agriculture-based crop selection analysis," IEEE Access,vol. 7, pp. 116965-116974, 2019.

[3] A. Suresh, N. Manjunathan, P. Rajesh, and E. Thangadurai, "Crop Yield Prediction Using Linear Support Vector Machine, "European Journal of Molecular & Clinical Medicine, vol. 7, no. 6, pp. 2189- 2195, 2020.

[4] M. Alagurajan, and C. Vijayakumaran, "ML Methods for Crop Yield Prediction and Estimation: An Exploration," InternationalJournal of Engineering and Advanced Technology, vol. 9 no. 3, 2020.

[5] P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Predicition of Crop Yeild Using SVM Approch with the Facility of E-MART System" Easychair 2020.

[6] S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomatograding and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, pp.103090, 2020.

[7] P.Tiwari, and P.Shukla, "Crop yield prediction by modified convolutional neural network and geographical indexes," International Journal of Computer Sciences and Engineering, vol. 6, no. 8, pp. 503-513, 2018.

[8] P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.

[9] N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," Ictact journal on softcomputing, vol. 11, no. 01, 2020

[10] A. A. Alif, I. F. Shukanya, and T. N. Afee, "Crop prediction based on geographical and climatic data using machine learning andDeep learning", Doctoral dissertation, BRAC University) 2018.

[11] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep Learning-based detector for real-time tomato plant diseases andpests' Recognition," Sensors, vol. 17, no. 9, pp. 2022, 2017.

[12] 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.

[13] K. A. Shastry, and H. A. Sanjay, "Hybrid prediction strategy to predict agricultural information," Applied Soft Computing, vol.98, pp. 106811, 2021.

[14] D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop Yield and Fertilizer Recommendation Using Machine LearningAlgorithms," International Journal of Engineering Applied Sciences and Technology, vol, no. 5, pp. 371-376, 2019.

[15] B. Devika, and B. Ananthi, "Analysis of crop yield prediction using data mining technique to predict annual yield of majorcrops," International Research Journal of Engineering and Technology, vol. 5, no.12, pp. 1460-1465, 2018.