JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CROP HARVEST PREDICTION AND DISEASE DETECTION USING MACHINE LEARNING

1st Dr. Poonam Sonar
Dept of Electronics and Telecommunication Engineering
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
poonam.sonar@mctrgit.ac.in

2nd Rajas Bhande
Dept of Electronics and Telecommunication Engineering
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
bhanderajas5@gmail.com

3rd Aditya Godambe
Dept of Electronics and Telecommunication Engineering
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
godambeaditya@gmail.com

4th Jayant Karale
Dept of Electronics and Telecommunication Engineering
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
jayantkarale0000@gmail.com

5th Rahul Padalkar
Dept of Electronics and Telecommunication Engineering
MCT's Rajiv Gandhi Institute of Technology, Mumbai, India
rahulpadalkar06@gmail.com

Abstract— **Crop prediction and disease detection are two important areas in agriculture that have received significant attention in recent years. Crop prediction involves forecasting the expected yield of a particular crop in a given season, which can help farmers to plan their activities and make informed decisions. Disease detection, on the other hand, it has identifiers and diagnoses diseases which affect the crop, which helps us to manage the disease. The prediction and detection process has been carried out mainly by machine learning algorithm and CNN. The research paper presents solution to above problem by developing a web page with the help of HTML, CSS , JavaScript for front end and Python and it's framework Flask for other aspects of the websites**

In this paper. We have proposed a website, which gives a system where a user can give his/her details of the soil, in the form fields as inputs. And then by using some machine learning algorithms and CNN modules. We recommend the best crop to grow in the soil. The Fertilizer which should be used to maintain the hygiene of the crop. And if the crop has some disease. Then with the help of different machine learning algorithms and modules, we detect the type of disease the crop has suffered from

Keywords*— *Crop, Crop Disease, Fertilizer, Prediction, Recommendation

I. INTRODUCTION

Agriculture is a vital sector that provides food, fiber, and other essential resources to sustain human life. The world has a growing population which will again increase to 9.7 billion by the year 2050, there is an increasing demand for agricultural products, which puts pressure on farmers to increase yields while also maintaining sustainability. In this context, crop prediction and disease detection have become crucial areas of focus in agriculture, this will eventually help the farmers to make the important decisions about the crop and there planning harvesting and management.

Crop prediction involves forecasting the expected yield of a particular crop in a given season, which is dependent on many factors such as weather patterns, soil quality, and historical data. Accurate crop predictions can help farmers plan their activities, optimize resource use, and reduce losses due to unexpected events such as droughts or floods. Disease detection, on the other hand, contains the identification of various diseases which may damage the crop, which can help farmers implement effective disease management strategies and prevent crop loss. The project is divided into three major parts : predicting crops , fertilizer prediction and leaf disease detection. Crop and fertilizer prediction is done with help of Random forest algorithm, and the crop disease detection is done by the CNN i.e the convolutional neural networks. The aim is to deploy a web page where user can interact with system and insert information like soil, temperature, humidity and the system will suggest the desired result. In the leaf disease detection the user will enter the leaf image and can identify disease of the crop.

II. LITERATURE REVIEW

Shujuan Zhang et al. [1] The research conducted by Shujuan Zhang in field of plant disease detection proposed a model to predict the diseases of crop plants in the early stages. The plant disease detection made use of the deep learning model . The deep learning proved very effective in their research as it provided with lot of flexibility and automatic feature extraction. The popular deep learning model CNN was used for image feature extraction and detection. The different plant diseases were included in the dataset and CNN was used for classifying the plant spot diseases and identification was done . The concept of hyper spectral imaging and CNN provided high accuracy results.

Nikola K Kasabov et al. [2] Nikola K. Kasabov developed a project for a crop prediction model with a different approach . The proposed solution by their paper presented concept of spiking neural networks which is used for the remote sensing and analysis of spatiotemporal analysis of image time series. The methodology included the analysis of the image time series and spiking neural network algorithm. The research and dataset was carried out with help of historical data of the crop yield. The remote sensing sensors were used and technical device called spectroradiometer was used for imaging purpose. The IOT implementation was stronger here. Rather than Web.

Potnuru Sai Nishant et al. [3] Potnuru Sai Nishant aim to develop a crop prediction model which would help farmers in producing better yield of crops. The idea behind the project was to develop a web based app for the farmers . The web based system was developed so that it would be easy for the farmers to understand and also the use machine learning models gave high accuracy results. The author used machine learning models like Kernel Ridge , Lasso and ENet algorithms to predict the yield of the crop and they also included concept of Stacking Regression. The different parameters considered were soil nutrients , which state the crop belongs to and the city. The web application was built by the team which gave result with high accuracy with help of stack regression model.

Dr. V. Geetha et al [4] The authors of this article focus on contract recommendation using the Random Forest algorithm. The Random Forest algorithm does more work, increasing the accuracy of the output and the speed of testing. The main purpose of this article is to review some authors and influence their planning processes to achieve better results. The Random Forest algorithm has many unique trees and they work like an army. Each unique decision tree in the random forest classifier is divided into estimator classes, and the class with the most votes becomes the estimator of the three species.

P.S Vijayabaskar et al. [5] In this paper, the author focuses on research on agricultural analysis of organic and inorganic agriculture, planting times of plants, income and loss information, and land management tools in only one region. An algorithm for unknown models that uses information provided by a set of isolated models. Manages organic, inorganic and real estate datasets. Estimates agricultural production from organic, inorganic and soil property datasets. All three are used in the algorithm and the planting is estimated by the multiple horizontal, predict and decision tree algorithm, respectively. Assists farmers with training, crop analysis; Provides weather forecast and also information about the state of agriculture in agriculture. In the previous process, many recycling methods were used from the existing data. During the planning process, the current price is used to estimate the future price that can be used to assess the ratio of arable and arable land for commercial real estate. The validity of the answers given to the questions.

M. G., Arunkumar et al. [6] Agricultural scientists around the world say a better machine is needed to predict and improve crop yields. The agricultural community strongly believes that proper crop management is essential for proper crop growth management. The difficulty of predicting crop yield is high due to the inadequacy of many different metrics and forecasting models that lead to crop loss. This research paper presents a model for product prediction.
Awan et al. [7] This article presents software development for estimating crop yields from weather and crop data. At the heart of the system is an unsupervised data partitioning method that uses kernel methods to find spatio-temporal patterns in weather data and provides the ability to manipulate complex data. For this purpose, a weighted kernel k-word algorithm with spatial constraints is proposed. The algorithm can check for noise, outliers and autocorrelations in the spatial data to analyze the data well, so oil palm can be predicted by analyzing the various factors that affect the results.

Bishnoy R.tos. [8] The ARMA and SARIMA models are the most accurate for national precipitation and temperature forecasting. Based on this prediction, crop yield is estimated using fuzzy logic method. It is based on information on areas such as temperature, humidity and soil pH.

S. V. Thombare et al. [9] The Indian economy is predominantly based on agriculture and the agro-processing industry. Soil composition (such as nitrogen, phosphorus, potassium), crop rotation, soil moisture, atmosphere and surface temperature, precipitation, etc. plays an important role in production. Current evidence in this area includes models for determining crop quality combined with machine learning algorithms (uniform forests, decision trees, neural networks). In this article, a proposed model was developed using deep learning techniques and accurate information was obtained about the price and value of the land needed, as well as crop estimation. It provides better accuracy than existing models. It analyzes the given data and helps the farmers to predict the crops which in turn helps in making money. To estimate profits, consider the climate and soil of the region.
The aim is to demonstrate a python-based system that uses decision rules to predict the best crop at the lowest cost in the conditions. In this study, SVM as machine learning algorithm, LSTM and RNN as deep learning method were used.

Gandge, Y et al. [10] Crop forecasting is the process of estimating crop yields using various variables such as precipitation, temperature, fertilizers, pesticides, and other atmospheric conditions and parameters. The process is very popular not only in agriculture. Data mining techniques are used and evaluated in agriculture to predict next year's crop. This article provides a brief summary of crop forecasting in selected regions of Mangalore, Kasargod, Hassan, Kodagu, India using the KNN algorithm.

III. PROPOSED METHODOLOGY

The crop prediction and disease detection system will utilize machine learning algorithms like Random forest algorithm , SVM , K means clustering and Convolution neural network. The website is built with the help of front end tools like

Bootstrap, HTML, Javascript, CSS. The backend of the website is built with help of Python , Flask and Jupyter notebook . The user will interact with the web page and can input the conditions related to his/her crop. The project consists of 3 parts that are crop prediction, fertilizer recommendation, disease prediction. Different conditions like the soil, temperature , humidity , ph. will be taken as input and the system will predict which crop should be planted by the user. Using the k, P, N (high and low) values we give a HTML recommendation to maintain the hygiene of the soil. And when the user uploads a photo of the crop then he/she is provided with the disease name of the plant if it has any.

1. Crop prediction

The main goal here is to ensure that people use the best products they have to grow on their farm. Results are based on several parameters and help farmers make informed decisions before planting. The data used in this project includes weather, soil etc. available in India. obtained from various data such as Unlike complex traits that affect crop yields, this information is simple yet very useful with a few trends. Information Nitrogen, Phosphorus, Potassium, pH, Precipitation, State, City etc. includes.The flow can be seen in Figure 1.

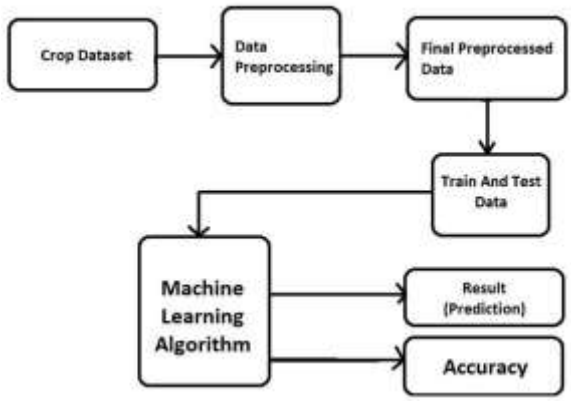

Figure. 1. System flow of the proposed method

A. Preprocessing

Preliminary data is the first step in machine learning design. Preprocessing involves collecting information about crop characteristics and correlating these characteristics with environmental variables. This information may include information such as crop type, planting date, weather conditions, land area and fertilization history. After collecting the data, preprocessing including cleaning, normalization and feature extraction should be done to make the data consistent and good.

B. Dataset Splitting

Data classification includes data classification found in training, validation, and test files. The main purpose of data classification is to evaluate the performance of the model on invisible data and to prevent overloading. Here the data is split in a ratio of 80:20, 80% of the data is chosen as reference data and 20% is also divided into two equal parts. Specifically, 10% of the data is sent to the test setup and 10% to the validation process.

C. Dataset training and testing

The training dataset for a crop recommendation system has a number of samples, each of which contains information about the soil properties, climate conditions, and the crops grown in that area in the past. This data is usually collected by agricultural experts and researchers, who analyze various factors such as soil type, pH level, nutrient content, temperature, rainfall, etc., to determine which crops are most suitable for a particular area After training the machine learning model on the training data, it is important to test its performance on the classified data to ensure that it performs well on new information. Test data is used to test the performance of the model on new and unseen data.

D. Machine learning algorithms

There are many kinds of algorithms used in the field of machine learning. But for this project we have used mainly 5 types of machine learning algorithms. Each algorithm has its own set of rules and procedure to give the required output. Different machine learning algorithms used in this project to predict the crop are Random Forest algorithm. Naïve Bayesian Classifier, K means Algorithm, SVM, XGBoost .The Table no. 1. shows the accuracy of each algorithm to predict the crop. We can see that the random forest algorithm has the best accuracy among all. While the support vector machine algorithm has the least accuracy.

Crop Prediction	
Algorithm	Accuracy
The tree of decision	0.917584241325689
Naïve Bayes	0.990909090909092
SVM	0.10681818181818183
Log Reg.	0.9522727272727274
Rand()Forest	0.990909090909097
XG Boost	0.9931818181818183

Table 1. : Crop Prediction

2. Fertilizer suggestion

In agriculture, fertilizers are an important part of increasing crop productivity. However, overuse of fertilizer can cause environmental problems,

including soil acidification and eutrophication. To optimize fertilizer use, machine learning algorithms can be used to estimate the amount of

Fertilizer Prediction	
Values	Description
K (High)	1. Dig the soil deeply with the shovel and water thoroughly to dissolve the potassium dissolved in the water. Allow the soil to dry completely and repeat digging and watering two or three times. 2. Sift the soil and use a soil sieve to remove as many stones as possible. Minerals found in rocks such as mica and feldspar slowly release potassium from the air into the soil.
K(Low)	1. Mix Potash or Potassium Sulphate 2. Try Seaweed Powder or Seaweed 3. Try Sul-Po-Mag
P (High)	1. Do not add fertilizers that are rich in soil nutrients but often contain too much phosphorus. 2. Plant nitrogen-containing vegetables (like beans and peas) to avoid phosphorus.
P (Low)	1. Add bone meal, a fast-acting source made from ground animal bones rich in phosphorus. 2. Phosphate Fertilizer Use a fertilizer high in nitrogen, phosphorus and potassium (for example: 10-20-10 is 20 percent phosphorus).
N (High)	1. Please fertilize abundantly, this is one of the easiest ways to level the soil with nitrogen fertilizer. 2. Plant nitrogen-fixing plants such as beans, beans, and soybeans from the legume family that reduce nitrogen in the soil.
N (Low)	1. Add sawdust or shavings to the soil The carbon in sawdust/sawdust loves nitrogen and helps absorb and absorb excess nitrogen. 2. Nitrogen-eating plants Tomatoes, corn, broccoli, cabbage, and spinach are examples of plants that thrive on nitrogen and absorb it. 3.Wetting your soil with water helps nitrogen sink deeper into the soil, reducing your plant's use.

Fertilizer suggestion table

fertilizer required for certain crops and soils. This article explores machine learning algorithms for fertilizer prediction, including decision trees, random forests, support vector machines, and neural networks. The data used in this study includes soil characteristics, climate data and crop data.The results show that the machine learning algorithm can predict the optimum amount of fertilizer and soils needed for a given crop, thereby increasing yields and reducing environmental impact . Farmers and agriculturists can use this method to make decisions about planting and ultimately can develop sustainable agriculture. Using machine learning for fertilizer forecasting can increase crop yields. Here, , which we consider as K, P, N value, is recommended to be fertilized.

HTML based fertilizer recommendation :

The table below shows the different description for the different values of K, P, N both High as well as low.

3. Disease detection :

Crop diseases can seriously affect crop yield and quality, causing economic losses for farmers and food shortages for consumers. Early detection and diagnosis of crop diseases is important to prevent their spread and minimize damage. Machine learning algorithms have demonstrated potential for crop disease prediction by analyzing large datasets of crop images and associated metadata. This article explores the use of machine learning algorithms such as convolutional neural networks (CNNs) in crop disease prediction. The data used for this study includes crop images from various databases and other government agencies.
The results show that machine learning can predict the efficacy of crops afflicted with a particular disease, allowing for early detection and improved disease management. Farmers and agronomists can use this method to make informed decisions about disease prevention and control, ultimately increasing yields and reducing bankruptcy.
Here we will use ResNet, each layer feeds into the next layer, we use a network with residuals, each layer to the next layer and directly to the layer about 2-3 hops away, to avoid overfitting (a case. when validation is done) the loss does not decrease at a certain point and more then it continues to increase while the learning loss continues to decrease). Figure 1. Shows the hierarchical structure of ResNet.

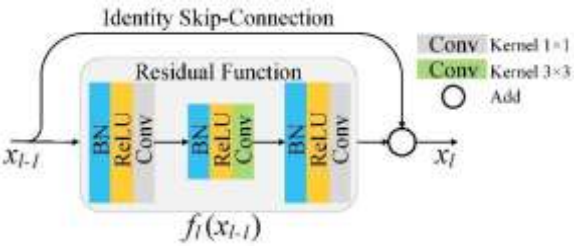

Figure. 1: Block diagram of ResNet

IV. EXPERIMENTAL SETUP

1. Crop Recommendation

A. Details of the dataset

The dataset we have collected is from Kaggle, which was the foundational need of the project. It consists of different types of columns such as Nitrogen, Phosphorous, potassium, pH, humidity and temperature.

B. Accuracy

Because we use a lot of machine learning algorithms to predict the crop. So each algorithm gives different sensitivity. Many machine learning algorithms for product prediction are Decision Trees, Naive Bayes, Support Vector Machine, Logistic Regression, Random Forest, XGBoost. Among all these algorithms, Random Forest has the highest accuracy and Support Vector Machine (SVM) has the lowest accuracy.

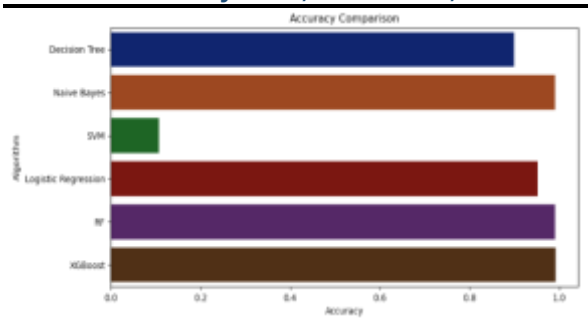

Figure. 2. Accuracy Comparision

C. Confusion matrix

Confusion Matrix is a method which is used to describe how well a categorization method performs, a confusion matrix is utilized. A confusion matrix summarizes and depicts how well a classification method performed. We may assess recall, accuracy, and precision using this helpful method

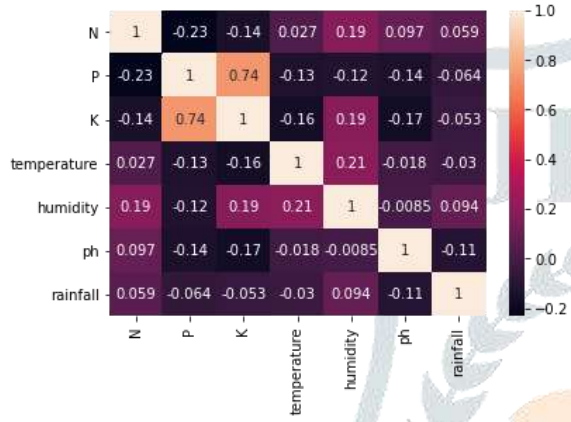

Figure. 3. Confusion matrix

2. Fertilizer suggestion

- Details of the dataset

In this dataset we have taken total 22 crop samples with different parameters. The parameters include Nitrogen, Phosphorous, Potassium, pH, soil moisture and the name of the crop. A user can select a crop from the list of 22 items. Each of the crop is mapped to a specific number.
Crops include rice, corn, chickpeas, kidney beans, pigeon peas, moth beans, mung beans, black beans, lentils, pomegranates, bananas, mangoes, grapes, watermelon, melons, apples, oranges, papayas, coconut, cotton, jute, coffee is found.

3. Disease Detection

A. Details of the dataset

The database contains approximately 87K rgb images of healthy and diseased crops, divided into 38 different classes. All data are split into 80/20 training and productivity keeping the sample logged. Next, create a new list with 33 test images for the prediction.

B. Accuracy

As the number of epochs increase, then simultaneously the accuracy also increases. The more we increase the epochs the more is the accuracy. So we can say that Accuracy is directly proportional to the number of epochs. The figure 4. Shows the graph of number of epochs vs the accuracy

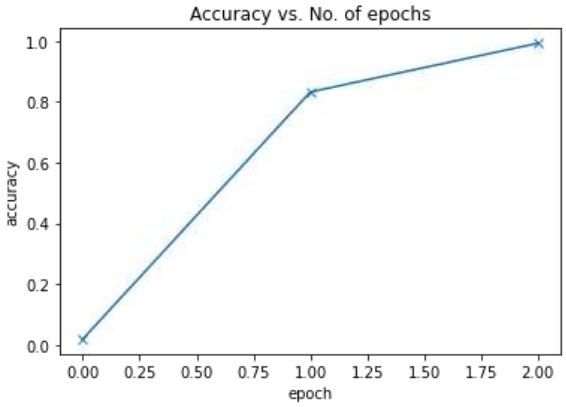

Figure 4. : Accuracy vs no. of epochs

V. CONCLUSION

The crop prediction and disease detection system is built with help of various machine learning algorithms such as Random Forest and convolution neural network. By analyzing the useful data about different environmental factors a predictive model is built to help farmers boost their crop growth. The algorithms such as Random forest and CNN are very powerful algorithms with high accuracy rates which has helped in these predictive models immensely. The ability to handle high dimensional data and automatic feature extractions make it very powerful and useful models. Overall with the use of these algorithms and implementing a standard methodology a website is built which will help farmers to improve crop productivity and crop security.

VI. ACKNOWLEDGEMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal and Dr. S. D. Deshmukh, Head of Department of Electronics and Telecommunication Engineering at the MCT's Rajiv Gandhi Institute of Technology, for providing us with the opportunity to work on our project. "Crop Harvest Prediction and Disease Detection Using Machine Learning". This project would not have been possible without the guidance and encouragement of our project guide, Dr. Poonam Sonar. We would also like to thank our colleagues and friends who helped us in completing this project successfully.

VII. FUTURE SCOPE

The future scope of the project is that we can develop an application for the farmers with real time features like weather conditions , soil prediction and yield prediction. Login and real time database can be added to monitor the different data of farmers. The login feature can be added which will help the farmer to keep the data of them saved. Which can be later used for reference purposes. The login feature will also greatly help the farmer to keep there data safe and up to them. The linking of the database will be beneficial to keep a record of all the user data that are visiting the website.

VIII. REFERENCES

[1] Lili Li, Shujuan Zhang, Bin Wang, "Plant Disease Detection And Classification By Deep Learning" in IEEE Access, China, 2021
[2] Pritam Bose, Nikola K. Kasabov, Lorenzo Bruzone,

Reggio N. Hartono, "Spiking Neural Networks For Crop Yeild Estimation Based On SpatioTemporal Analysis Of Image Time Series", in IEEE Transactions on Gio Science and Remote Sensing, 2016

[3] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber, "Crop Yeild Prdiction Based on Indian Agriculture using Machine Learning", in International conference for Emerging Technology(INCET), 2020

[4] Dr. V. Geetha, A. Punitha, M. Abarna, m. Akshaya, S. Illakiya, AP. Janani, "An Effective Crop Prediction using Random Forest Algorithm" in International Conference on System, Computation, Automation and Networking (ICSCAN), 2020

[5] P.S. Vijayabaskar, Sreemathi.R.Keertanaa.E, "Crop Prediction using Predictive Analytics, in International Conference on Computation of power, Energy, information and Communication,(ICCPEIC), 2017

[6] Ananthara, M. G., Arunkumar, T., & Hemavathy, R. (2013, February). CRY—an improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 473-478). IEEE

[7] Awan, A. M., & Sap, M. N. M. (2006, April). An intelligent system based on kernel methods for crop yield prediction. In Pacific-Asia Conference on Knowledge Dis

[8] Bang, S., Bishnoi R.., Chauhan, A. S., Dixit, A. K., & Chawla, I. (2019, August). Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models. In 2019 Twelfth International Conference on Contemporary Computing (IC3) (pp. 16). IEEE.

[9] Bhosale, S. V. Thombare, R. A., Dhemey, P. G., & Chaudhari, A. N. (2018, August). Crop Yield Prediction Using Data Analytics and Hybrid Approach. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-5). IEEE

[10] Gandge, Y. (2017, December). A study on various data mining techniques for crop yield prediction. In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT) (pp. 420-423). IEEE.