Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Forecasting Crop Production Based On Soil Data and Detection of Diseases Using Machine Learning

M. Mohammed Riyaz[1], V. Gopalakrishnan[2], S. Kalidasan[3]

[1,3]MCA, Thanthai Periyar Government Institute of Technology, Anna University, Vellore, India
[2]EEE, Government College of Technology, Anna University, Coimbatore, India

E-mail: [1]mbajju@gmail.com, [2]gopalgct@hotmail.com, [3]kalidasanvellore@gmail.com

Abstract

Estimating crop creation has been a consistent worry since the start of the historical backdrop of agribusiness. Guaging strategies have advanced, as has horticulture itself and the determinations of the gauges required. The individuals who utilize figure information are continuously looking for more noteworthy precision, granularity, likeness, and idealness. The people who produce the information or add to their creation generally work under monetary and specialized requirements. Acquiring opportune information presents an undeniable test. Today, the human, institutional, specialized, and monetary foundation behind crop figures and proposal gauges specifically can be unquestionably complicated. This distribution gives experiences into such complex information frameworks at the nation level. It features great practices and prospects for what's to come. The distinguishing proof and location of sicknesses of plants is one of the central matters which decide the deficiency of the suggestion of yield creation and agribusiness. The investigations of plant illness are the examination of any noticeable places in any part of the plant, such as any spots or conceals, which assists in distinguishing between two plants. The manageability of the plant is one of the central issues for rural turn of events. The discovery of plant sicknesses is undeniably challenging to get right. The recognizable proof of the illness requires loads of work and ability, bunches of information in the field of plants and the investigations of the identification of those sicknesses. Thus, picture handling is utilized for the identification of plant infections. The detection of infections follows the strategies for picture procurement, picture extraction, picture division, and picture pre-handling.

Keywords: Machine Learning, crop production, image pre-processing, crop recommendation prediction

DOI: https://doi.org/10.36548/jscp.2022.1.003
Received: 16.03.2022, received in revised form: 05.04.2022, accepted: 20.04.2022, published: 28.04.2022

1. Introduction

A Machine Learning (ML) manages issues where the connection among info and result factors isn't known or difficult to acquire. The "learning" term here indicates the programmed procurement of underlying depictions from instances of what is being portrayed. Not at all like conventional measurable techniques, doesn't ML make suspicions about the right design of the information model, which depicts the information. This trademark is extremely valuable to demonstrate complex non-straight ways of behaving, like a capacity for crop suggestion forecast.

ML strategies generally effectively applied to Crop Recommendation Prediction (CYP). Managed learning calculation comprises of an objective/result variable (or ward variable) which is to be anticipated from a given arrangement of indicators (free factors). Utilizing this arrangement of factors, a capacity that guides contributions to wanted yields is created. The preparation cycle goes on until the model accomplishes an ideal degree of precision on the preparation information.

1.1 Overview of the system

The previous situation for plant infection location included direct eye perception, recollecting the specific arrangement of illness according to the environment, season and so on. These strategies were for sure mistaken and very tedious. The flow strategies for plant infection identification included different research center tests, talented individuals, exceptional labs and so forth. These things are not accessible wherever particularly in far off regions.

Location of infection through some programmed method is useful in light of the fact that it lessens a curiously large work of watching in immense homesteads of harvests, and at frightfully beginning phase itself, it distinguishes the side effects of illnesses intends that after they appear to be on plant leaves. There are multiple ways of recognizing plant pathologies. A few illnesses have no apparent side effects, or the impact becomes observable beyond any good time to act, and in those circumstances, a modern examination is required.

2. Related Work

Santosh Mahagaonkar et al. (2019) [1] demonstrated that expectation of harvest yield in light of area and legitimate execution of Algorithms, accomplishes the higher harvest

yield. From above work it is inferred that for soil arrangement Random Forest was great with exactness 86.35% contrast with Support Vector Machine. For crop yield expectation Support Vector Machine was great with precision 99.47% contrast with Random Forest Algorithms.

Yemeserach Mekonnen et al. (2020) [2] carried out an IoT based information driven model for an incorporated food, water, and energy framework. The principle objective was to screen and gauge the three reliant assets utilizing remote sensor organizations and IoT stage across the entire framework. To effortlessly explore the information procurement and coordination, Farming portable application is planned and executed with Google Firebase distributed storage as a back-end.

3. Proposed Work

In the proposed framework, a camera picture or a video feed is needed to handle the picture, utilizing PC vision with open cv library handling the picture will be the main thing, the picture will be handled and changed over to number information in a scope of RGBI. Crop Selection to boost the harvest suggestion, choice of the suitable yield that will be planted assumes an essential part. It relies upon different elements like the sort of soil and its arrangement, environment, geology of the area, crop suggestion, market costs and so forth.

Procedures like Artificial neural networks, K-nearest neighbors and Decision have cut a specialty for themselves with regards to trim determination which depends on different variables. Crop determination in view of the impact of normal catastrophes like starvations has been done in light of machine learning. The utilization of fake brain organizations to pick the harvests in light of soil and environment.

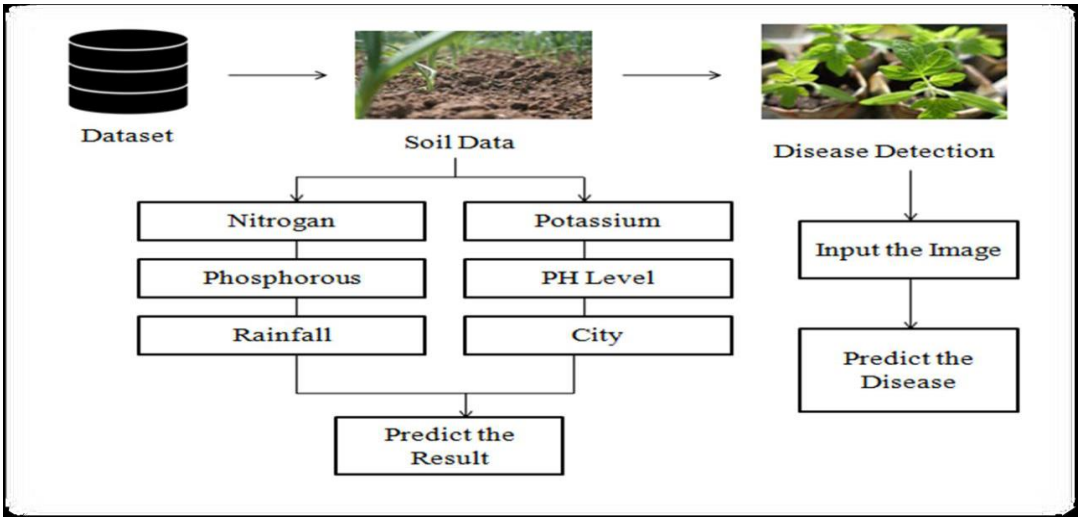

Figure 1. System architecture

A plant supplement the executive's framework has been proposed in view of machine learning methods to address the issues of soil, keep up with its fruitfulness levels, and thus further develop the yield suggestion. A selection technique called CSM has been proposed which helps in crop determination in light of its suggestion forecast and different variables.

3.1 Dataset

Generally researchers used .csv files of agriculture dataset for crop recommendation prediction. The dataset is supervised learning. It comprises of various qualities like County Name, State, dampness, temperature, NDVI, wind Speed, suggestion and so forth. Plant Village dataset consists of 55,780 images of different plant leaves which are divided into 18 classes. The dataset consists of 16 types of plant species and 36 types of plant diseases.

The dataset contains both healthy and diseased crop images. The images cover 12 species of crops, including: apple, blueberry, grape, orange, peach, pepper, potato, raspberry, soy, strawberry, and tomato.

Each class comprises of two fields for example name of the plant and name of the sicknesses. The soy bean leaf is considered for characterization. Every one of the images are resized and fragmented for pre-handling and further characterization.

3.2 Polynomial regression

In this model regression investigation is finished by inferring the connection between x autonomous variable and y subordinate variable as most extreme limit polynomial of x. A non-straight relationship is fitted to regression model x and mean y, which is signified by E(y | x), and is utilized for peculiarity resembles development pace of tissues, residue of carbon isotopes in lakes, and scourges infection movement.

Regardless of the way that polynomial relapse fits a non-straight model, as a measurable assessment it is considered as direct, as a capacity E(y | x) as obscure boundaries from the information. Because of this polynomial relapse is viewed as unique instance of different direct regression.

3.3 Decision Tree

As decision tree utilizes greedy Technique, trait picked in the initial step can't be utilized later to give better arrangement of information. If at all it is utilized in the subsequent stages Decision Tree over fits the preparation dataset that can prompt unfortunate outcomes.

Troupe model is fused to beat this downside and promising outcomes are gotten by group models.

3.4 Image (RGB) load

The images of the plant leaf are captured through the camera, this image is in RGB (Red, Green and Blue) structure, variety change structure for the leaf picture is made, and afterward an autonomous variety space change for the variety change structure is applied.

3.5 Pre-processing

To eliminate noise in image or other object evacuation, it is considered to preprocessing methods. Image cutting for example editing of the leaf image to get the intrigued image area. Image smoothing is finished utilizing the smoothing channel. Image improvement is completed for expanding the difference.

The RGB images into the dim image utilizing variety change utilizing equation(x) =0.2990*R + 0.5871*G + 0.115*B Then the histogram equalization which disseminates the forces of the image is applied on the image to improve the plant illness image. The aggregate circulation work is utilized to disseminate force values.

3.6 Segmentation

Division of leaf image is significant while handling image from that Segmentation implies dividing of image into different piece of same elements or having some comparability. The division should be possible utilizing different techniques like k-means clustering.

3.7 Pre-processing of the Images

Pre-processing is a vital stage in CNN as the images in the dataset may have some irregularity which might influence the precision of the framework. The images in the dataset have noise and non-uniform lighting which should be amended in this progression. Division is applied on the images to dispose of lopsided foundations.

Through division, the pertinent piece of the images is removed which for this situation are the image of leaves. Subsequently, after division the images of leaves with black background is obtained. Presently to correct the non-uniform lighting, the images are converted to grayscale images and send for further processing.

3.8 Feature Extraction

As the grayscale images are obtained from the previous step, the images are converted into reduced variables. Basically each pixel of the images are taken and are converted into matrix for performing convolutions. The process runs across all the pixels where the convolution matrix is simply multiplied with each pixel matrix.

3.9 Disease detection and classification

Detection of disease is performed in two steps i.e., detection of the type of crop and detection of type of disease. This takes place with the help of Convolutional Neural Network. Transfer Learning is used for building the Model. When the leaf is healthy and there is no classification, and the results are shown as healthy and when there is a disease which when grey scaled shows black spots, it classifies them so they are shown as which disease they are and the confidence of the classification. Classification takes place between two numerical arrays. If the numerical arrays match, then it is a healthy or a diseased leaf, depending upon the dataset provided. Classification is a simple but relevant procedure which gives a proper result and is used in plant disease identification.

3.10 Features

The highlights of the plants range from color, shape and disease type. This can make a million odd diseases which the model has to characterize and put in the system. As the features are put into various categories the images are decoded to get the next result, which is the layered output.

3.11 Layer Outputs

The layer outputs are different layers of the detection of the disease. These can be the grayscale images and the rgb of the images. These help in separating different colored features of the leaves and make sure they are categorized into multiple different categories. It also makes sure that the user can understand where the disease is located, and the numerical value can easily make the white and black images into 1 and 0.

3.12 Prescient ability

The precision of the different methodologies in anticipating proposals in years that were not used to prepare the model has been considered. The exactness of the parametric model and the SNN was considerably improved by stowing, however the packed away SNN

performed best. The completely nonparametric brain net which was prepared indistinguishably from the SNN yet needed parametric terms performed significantly more terrible then either the OLS relapse or the SNN. That sacking works on model attack of both the OLS relapse and the SNN infers that specific years might have filled in as factual influence focuses, and as with the end goal that un-stowed suggestion models may over fit the information.

This is on the grounds that there are too barely any unmistakable long stretches of information to decide if the fieriness of a bizarrely hot year is truth be told the reason for that year's abnormally low suggestions. On the off chance that bootstrap tests that preclude such years gauge various connections, averaging such gauges will diminish the impact of such anomalies

4. Results and Discussion

Prerequisites based testing is a trying methodology wherein experiments, conditions and information are gotten from necessities. It incorporates useful tests and furthermore non-practical properties like execution, dependability or ease of use. Stages in Requirements based Testing:

- Characterizing Test Completion Criteria-Testing is finished just when all the practical and non-utilitarian testing is finished.
- Configuration Test Cases-A Test case has five boundaries specifically the underlying state or precondition, information arrangement, the data sources, anticipated results and genuine results.
- Execute Tests-Execute the experiments against the framework under test and archive the outcomes.
- Check Test Results-Verify assuming the normal and genuine outcomes match one another.
- Check Test Coverage-Verify assuming the tests cover both utilitarian and non-useful parts of the necessity.
- Track and Manage Defects-Any imperfections distinguished during the testing system carries on with the deformity life cycle and are followed to goal. Deformity Statistics are kept up with which will provide us with the general status of the task.

5. Conclusion

The results show that, an exact crop suggestion forecast can be accomplished utilizing the Random Forest Algorithm. The algorithm accomplishes a biggest number of crop proposal models with most reduced models. It is appropriate for monstrous crop suggestion expectation in agricultural planning. This makes the farmers to take the ideal choice for right yield to such an extent that the farming area will be created by inventive thoughts.

References

[1] S. S. Sannakki and V. S. Rajpurohit," Classification of Pomegranate Diseases Based on Back Propagation Neural Network," International Research Journal of Engineering and Technology (IRJET), Vol2 Issue: 02 | May-2015

[2] P. R. Rothe and R. V. Kshirsagar," Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC),2015.

[3] Aakanksha Rastogi, Ritika Arora and Shanu Sharma," Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic" 2nd International Conference on Signal Processing and Integrated Networks (SPIN)2015.

[4] Godliver Owomugisha, John A. Quinn, Ernest Mwebaze and James Lwasa," Automated Vision-Based Diagnosis of Banana Bacterial Wilt Diseaseand Black Sigatoka Disease ", Preceding of the 1'st international conference on the use of mobile ICT in Africa, 2014.

[5] uan Tian, Chunjiang Zhao, Shenglian Lu and Xinyu Guo," SVM-based Multiple Classifier System for Recognition of Wheat Leaf Diseases,"Proceedings of 2010 Conference on Dependable Computing (CDC'2010), November 20-22, 2010.

[6] S. Yun, W. Xianfeng, Z. Shanwen, and Z. Chuanlei, "Pnn based crop disease recognition with leaf image features and meteorological data," International Journal ofAgricultural and Biological Engineering, vol. 8, no. 4, p. 60, 2015.

[7] J. G. A. Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," Springer Plus, vol. 2, no.660, pp. 1–12, 2013.

[8] Caglayan, A., Guclu, O., & Can, A. B. (2013, September). "A plant recognition approach using shape and color features in leaf images." In International Conference on Image Analysis and Processing (pp. 161-170). Springer, Berlin, Heidelberg.

Author's biography

M. Mohammed Riyaz received MCA degree from Anna University in 2012. After that, joined Thanthai Periyar Government Institute of Technology as an Assistant Professor. He is currently pursuing Ph.D. degrees in Anna University His research interests include Block chain, Internet of Things, and Machine learning.

V. Gopalakrishan Professor &Head of the Department & Vice Principal is currently a Professor of EEE in Government College of Technology, Coimbatore-13. His research is in the broad areas of Computer Network and Internet Research. He has published over 13 Journals and have participated in 16 National and International Conferences.

S. Kalidasan received an MCA degree from Anna University in 2009. After that, joined Thanthai Periyar Government Institute of Technology as an Assistant Professor. He has published 4 Journals. His examination advantages incorporate Big Data, Cloud Computing, Data Science and Machine learning.