



# INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact Factor: 6.078

(Volume 8, Issue 3 - V8I3-1271)

Available online at: <https://www.ijariit.com>

## Best Crop Yield Prediction and Maintenance of Soil Fertility using pH Value of Soil: Predicting pH Value using Digital Image Processing and Linear Regression

Himanshu Ajay Borade

[himanshu.21810267@viit.ac.in](mailto:himanshu.21810267@viit.ac.in)

Vishwakarma Institute of Information  
Technology, Pune, Maharashtra

Amogh Chandrashekhar Kurlekar

[amogh.21810077@viit.ac.in](mailto:amogh.21810077@viit.ac.in)

Vishwakarma Institute of Information  
Technology, Pune, Maharashtra

Pranjal Shailendra Sawarbandhe

[pranjal.21810265@viit.ac.in](mailto:pranjal.21810265@viit.ac.in)

Vishwakarma Institute of Information  
Technology, Pune, Maharashtra

Randhir Arun Gitte

[randhir.21810214@viit.ac.in](mailto:randhir.21810214@viit.ac.in)

Vishwakarma Institute of Information Technology, Pune, Maharashtra

Dr. Rohini Chavan

[rohini.chavan@viit.ac.in](mailto:rohini.chavan@viit.ac.in)

Vishwakarma Institute of Information Technology, Pune,  
Maharashtra

### ABSTRACT

*pH of soil is a key factor which can impact on nutrients in soil and plays an important role in deciding crop which will we yield in farm as well as fertilizer that farmer will use to make their crops better. The traditional methods are proven by many scientists and have their own disadvantages like sensor damaging and maintenance of kits. The alternate solutions such as image processing based prediction gives accurate and fast results than traditional methods. Here, we have used image processing and Linear Regression algorithm to identify the pH level of soil. Basically, to determine the pH level of soil, pH kits of labs are used. But they are not always accurate as they add extra maintenance for the results. To overcome this issue, we proposed a system that can obtain the pH value, recommend the crop for the same and also show nearest fertilizer stores to the farmer. For crop recommendation system it uses attributes like soil pH, texture of soil, colour of soil and nutrients available in soil. In this paper, we proposed a method to get the result of soil pH instantly, so that it will be easier to farmer to analyse the crop prediction.*

**Keywords:** Linear Regression, Agriculture, pH of the soil

### 1. INTRODUCTION

As we know about 60% of the Indian population is engaged in agriculture as a primary source of livelihood. In spite of this it contributes only 20% to the GDP of the country. This is because of lack of technologies. After availability of water, soil quality also plays an important role in production of crops. The nutrients present in the soil are also responsible for quality of crops. Different crops need different nutrient composition in the soil. So, soil is a vital source of nutrients required by plants to grow.

The three main nutrients in soil which are nitrogen, phosphorous and potassium generally called as NPK. Even the fertilizers are given in the composition of NPK. Basically, these nutrients are affected by the pH of soil. As a result, government has provided farmers with soil testing labs so that they can find out soil. So, they will use the fertilizers accordingly and will cultivate the crops which will be suitable.

Basically, farmers need to collect the soil sample and had to visit the government or private soil testing laboratories with their soil sample in order to check the nutrient composition in the soil. But by using this technique farmers will have a direct access to check the nutrient composition at their farm itself. We are using image processing to check the pH of the soil sample uploaded by the farmer. For this we will have a dataset of more than thousands of images. So, the result provided will be instant. Otherwise, farmers have to wait for 2-3 days or sometimes for a week to get the result from soil testing laboratories.

For the dataset we have used Realme XT mobile phone, having Sony f 1.9 apertures. All the soil samples were first collected, and then they were all taken to an indoor place to take photos. So that the lighting conditions should be same and there will be no

problem while removing the background. Also, the background for every soil sample was same. We try to take photo of every soil sample from same angle and position.

## 2. LITERATURE SURVEY

Agriculture is the oldest and traditional business of our country India. It is more like that every citizen is dependent on agriculture in our ancient times. The traditional method that experts use in the labs is bit longer to get result. Many labs are under the Ministry of Agriculture, Government of India while some are private. Government labs are much cost effective. On the other hand, private labs are much expensive.

For our current project, we referred various research paper based on soil testing and methodologies of soil testing. We mentioned all the papers that we referred in reference section of this paper.

[1] In this paper, researchers proposed a methodology for best crop selection. In this, they collected the data regarding the weather, NPK (Nitrogen, Phosphorous, and Potassium), pH of soil and based on this data, they predicted which crop will be suitable for that particular field of farm. They used classification in their particular paper.

[2] This paper was introduced us the digital image processing method for soil. In this paper, researchers used various algorithms such as Artificial Neural Network (ANN), K-Nearest Neighbour (KNN). The methodology was simple, they first created dataset with images of soil, masked them, extracted features and according the features they did various calculations. They predicted the value of pH using KNN first and ANN afterwards. Using KNN is better than ANN as they got accuracy of 84% while predicting the pH value.

[3] This research paper proposed a method to find out soil colour which will be used for analysing different components of soil. They have used a Classification technique to classify different soil of different colour.

[4] In this paper, the soil from Ramtek region is gathered and using CNN and features extraction they predicted the value of pH and fertilizer recommendation according to the need of farmer.

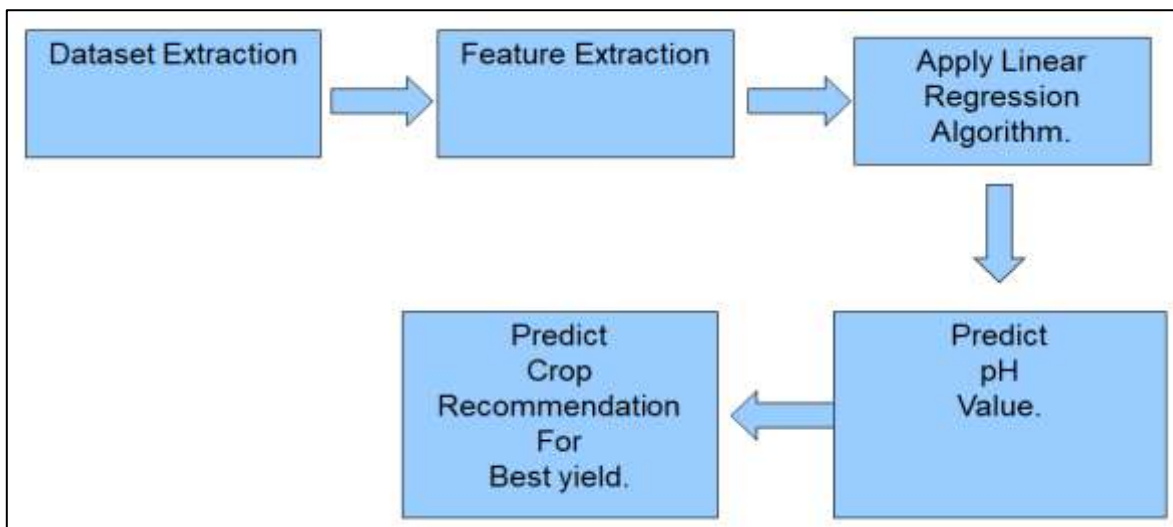
### Traditional Methods for Soil pH Testing

To test the soil pH, soil sample is taken to laboratories. But traditionally other methods were used. If test is done for checking alkalinity of soil, small amount of water is added to the soil and same amount of vinegar as that of water. If the soil sample is alkaline then bubbles will be produced with a hissing sound. More the bubbles, more alkaline is the soil i.e. higher the pH. When the more acidic vinegar comes in contact with something alkaline, bubbles are produced.

To test the acidity of soil, small amount of water with same amount of baking soda is added to the soil. If the soil is acidic, then there will be vigorous reaction. More vigorous the reaction more is the acidity i.e. lowers the pH. When the more alkaline baking soda comes in contact with the acidic soil, the reaction takes place.

In these traditional methods, the pH was not used to be accurate. We can just identify the acidity or alkalinity of the soil, but not the exact values that we can obtain in modern methods.

## 3. METHODOLOGY



### 3.1 Implementation plan of methodology for our project: Best Crop Yield prediction and Maintenance of soil fertility using pH value of soil: Predicting pH value using digital image processing and Linear Regression

The images of soil are captured by a mobile smartphone with its camera which has aperture f 1.9. All images are of high dimensions, in 8 bit- RGB format and saved in JPG format in Google drive. Then further these photos undergo in certain processes like dimension reduction, masking, pre-processing, augmentation.

### 3.1 Pre-processing

- We separated the dataset and pH of that particular image in excel file. In google colab, we are taking from both simultaneously.
- In openCV, generally images are stored in BGR .

```
open_cv_image = open_cv_image[:, :, ::-1].copy()
soiling.append(open_cv_image)
```

### 3.2 Dimension Reduction

- To get a stable dataset with same light and same dimension and same white background, to make resizable images is good choice.
- Hence, we resize image in 256 \* 256 dimensions. By doing this, we got the prefect dataset for this.

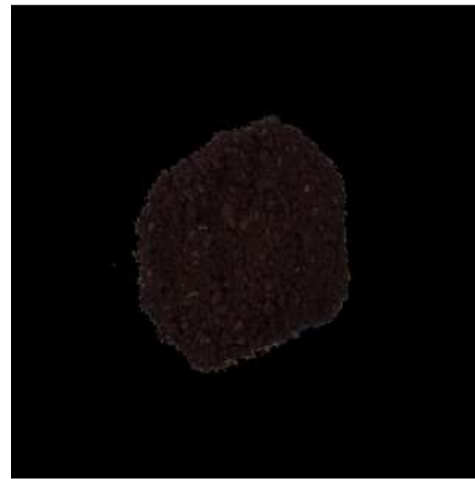
```
for i in range(len(soiling)):
    resized.append(cv2.resize(soiling[i], (256,256), interpolation = cv2.INTER_AREA))
```

### 3.3 Feature Extraction

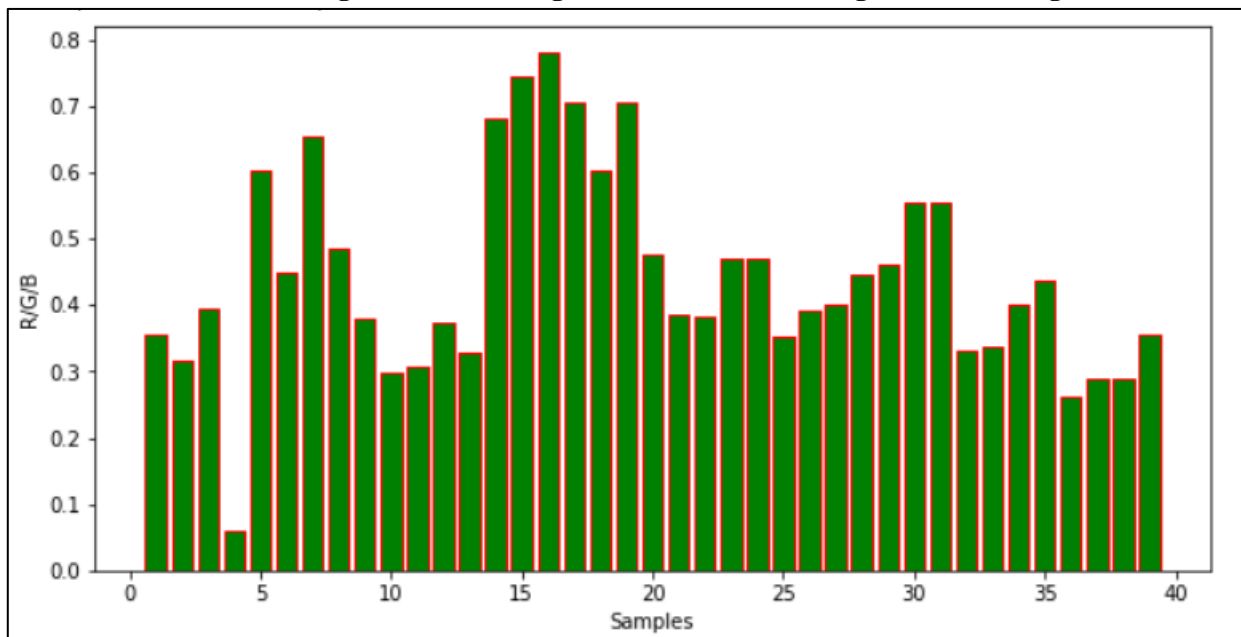
- For feature extraction, we did the masking. For this, as the brown images pixels are less than 90 and white images pixels greater than 90 We made pixels greater than 90 is equal to 0.
- After masking, getting the RGB features became easy. We extracted the values of RGB as features and plotted them.



3.2 Soil Image before masking



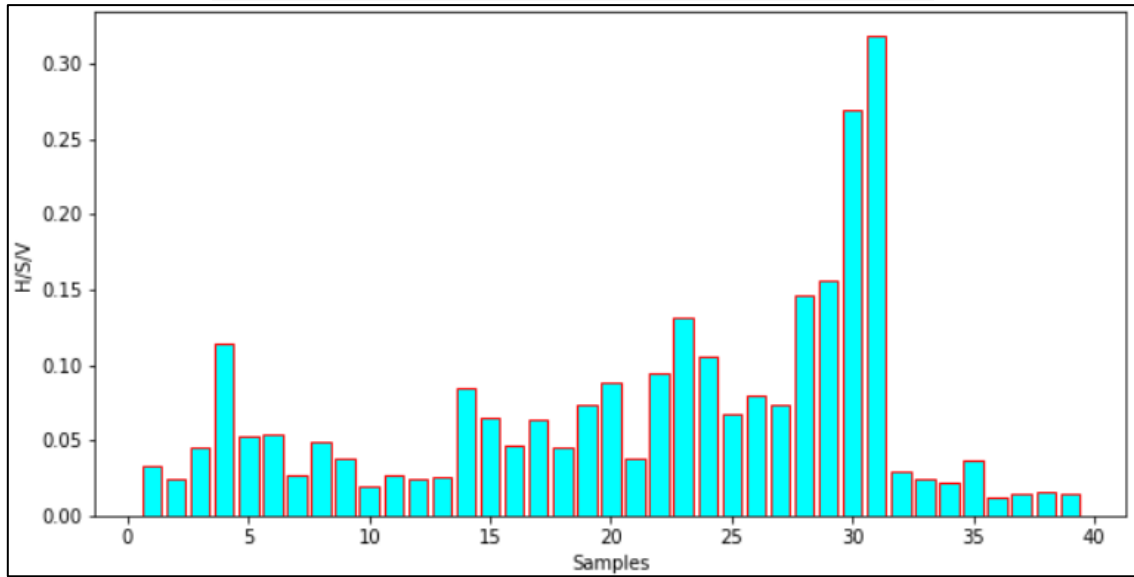
3.3 Soil Image after masking



3.4 R/G/B plotted against samples of soil

- Also, HSV (Hue Saturation Value) plays important role here. We found out these values and plotted them against the soil samples.

```
H.append(average_color[0])
S.append(average_color[1])
V.append(average_color[2])
divide=np.divide(H,S)
hsv=np.divide(divide,V)
```



**3.5 H/S/V plotted against samples of soil**

**3.4 Augmentation:**

- For making any dataset able to validate and test, we need to augment that data. Here we used image generator function to augment the images.

```
width_shift_range=.2,
height_shift_range=.2,
rotation_range=45,
horizontal_flip = True,
vertical_flip=True)
```

- And then we stored all augmented photos, extracted features in Google Drive. Hence, using it in our next part i.e. for linear regression.

**3.5 Main Algorithm:**

**Table 3.6 Features extracted from images of soil**

Mean_R	19.6370362222413
Mean_G	12.796971974355397
Mean_B	12.312919029822716
Mean_H	12.601635272686298
Mean_S	34.371478182055576
Mean_V	19.929198268568996
S1	Mean_saturation
S2	Mean Saturation+ Mean_Hue

The extracted features of images of soil are put together in table 3.6. According to [2], the S2 value i.e. mean saturation+ mean Hue value. Here S1 is a mean value of saturation. As it consists of array with all the pixels of saturation.

**4. RESULT AND DISCUSSION**

- Linear Regression: It is a machine learning algorithm which is used to predict any data of pH of soil when required dataset is given. Its coefficient gives the value of co-relation gives performance of regression. The regression value is good between 0.8-0.99. The Value between 0.8-0.99 is considered as best for algorithm.
- According to [2], the two matrixes which are termed as S1 and S2 in our research paper give higher co-relation value. We tested both of them in our algorithm and we got the same result as they are showing same high co-relation.

S1	Lower Co-relation	0.83412
S2	Higher Co-relation	0.86123

## **5. CONCLUSION**

pH of soil is a key factor which can impact on nutrients in soil and plays an important role in deciding crop which will we yield in farm as well as fertilizer that farmer will use to make their crops better. The traditional methods are proven by many scientists and have their own disadvantages like sensor damaging and maintenance of kits. The alternate solutions such as image processing based prediction gives accurate and fast results than traditional methods. Here, we have used image processing and Linear Regression algorithm to identify the pH level of soil. Here, using this algorithm we got accuracy of 86% , We input one image that we have tested in traditional algorithm as a result we got pH 8.1 and using this algorithm we got pH as 8.07. As we increase the dataset, the result will be much better and accuracy will also increase.

## **6. REFERENCES**

- [1] Mr. Rakesh Kumar, Mr. MP Singh, Mr. Prabhat Kumar , “Crop Selection method to maximize crop yeild rate using machine learning technique”, 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM) , 2015,pp 256-263.
- [2] Utpal Barman, Ridip Choudhary, Gauhati University, Assam, India, “Prediction of soil pH using smartphone based Digital Image processing and prediction algorithm”, Vol.-14, No.2, March-April (2019,) pp 226-249.
- [3] Shima Ramesh Maniyath, Mr. Ramchandra Hebbar, MVJ college of engineering, Bangalore, Karnataka. ,“Soil Colour Detection using KNN Classifier”, Conference: 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 2018.
- [4] D.Anantha Reddy, Bhagyashri Dadore, Aarti Watekar, Kavikulguru Institute of Technology and Science, Ramtek, Nagpur., “Crop recommendation system to maximize crop yield in Ramtek Region using Machine Learning”, International Journal of Scientific Research in Science and Technology, 2019.
- [5] <https://www.mpcb.gov.in/laboratory-analysis-charges/Soil-Sludge-Sediment-Solid-Waste>
- [6] <https://timesofindia.indiatimes.com/city/nagpur/illiteracy-main-cause-behind-farm-suicides/articleshow/62044463.cms>