Predication of soil pH using HSI colour image processing and regression over Guwahati, Assam, India

Utpal Barman*
Gauhati University, Guwahati- 781014 (Assam), India
Ridip Dev Choudhury
Gauhati University, Guwahati- 781014 (Assam), India
Niyar Talukdar
Girijananda Choudhury Institute of Management and Technology Azara, Guwahati-781017 (Assam), India
Prashant Deka
Girijananda Choudhury Institute of Management and Technology Azara, Guwahati-781017 (Assam), India
Indrajit Kalita
Girijananda Choudhury Institute of Management and Technology Azara, Guwahati-781017 (Assam), India
Naseem Rahman
Girijananda Choudhury Institute of Management and Technology Azara, Guwahati-781017 (Assam), India

*Corresponding author. E-mail: utpalbelsor@gmail.com

Abstract
Soil is known to be the most valuable natural source for all agriculture fields. Soil has two properties, namely- physical and chemical. These properties include soil moisture, texture, etc. and the latter include pH value. Soil texture plays an important role in crop cultivation. The physical properties of soil such as texture and granular size determine the water and nutrient holding capacity. Also the chemical property like pH value is very important for plant growth and development. Soil having pH value between 5.5 and 7 is optimal for agricultural purpose. Hence, a detailed study of soil pH property is necessary for cultivation. But laboratory method of soil pH calculation is a very costly and tedious process. Therefore, it is essential to develop an expert-based system that will overcome this issue. However, the system must be able to give correct result and should match with those conducted in laboratory. Farmers analyze pH either in lab or by soil pH card based on soil image color. But this is not an effective method since it relies heavily on human perception. Hence, we have developed an expert based system which can determine the pH of the soil without any human error. For this, we have conducted our experiments with the help of MatLab tool and smart phone as we have concerned about the rural farmers. We have analyzed and compared the proposed system results with the traditional laboratory methods with regression and have found 86 % accuracy in our model.

Keywords: Color Image Processing, Feature Extraction, pH, Regression, Soil

INTRODUCTION

Soil physical and chemical properties always play an important role in agriculture fields and also in some engineering activities. The physical properties of the soil mean soil texture, soil moisture and soil permeability. Soil texture indicates the shapes and size of the soil. The textures of the soil are sand, silt and clay. Soil moisture means the water holding capacity of the soil. Soil permeability is the property of soil which deals with the ability of water to flow through it. Here the water moves through the holes or voids in the soil and seeps through them. For examples, Sand is highly permeable and clay is least permeable. The chemical properties of the soil mean soil inorganic Compounds and soil pH. Many inorganic compounds maybe present in soil like potassium, magnesium, zinc which is always helpful for plant growth. Soil pH deals with the calculation or measure of acidic value or alkaline value of a soil. A soil with a pH of 6.0 is ten times more acidic than a soil of Ph 7.0. Changes in soil Ph dramatically affect the availability of nutrients to growing crops [Babu *et al.,* 2011].Among all these properties, soil pH and soil texture are the important properties of soil as their impact on cultivation is more as compare to other soil properties of soil. Soil texture primarily deter-

mines the water holding capacity of soil. For example, the water holding capacity of silt and clay is more as compare to sandy soil. Apart from this soil texture plays an important role in crop selection. Soil pH determines the acidity and alkaline properties of the soils. Rather than this, crop cultivation of an area depends on the quality of the soil. It means, soil having proper texture and pH value in between 5.5 to 7 is optimal for agricultural purpose. So it is important to determine the texture, moisture and pH properties of the soil. But all the laboratory methods of soil texture and pH calculation are time and labor consuming. For example, hydrometer method is used to find the percentage of sand, silt and clay of the soil. Based on these values, soil texture classification is done. The hydrometer method has taken more than 24 hours for the procedure. Like hydrometer, oven dried method is used to find the soil moisture and the whole procedure has taken more than 24 hours for the soil samples.

 In laboratory, soil pH is calculated with the help of soil pH meter and soil color pH card. The soil pH meter procedure has taken more than 1 hour for a simple soil sample. With the advance of technology and use of computers automation is being used more in day to day life. It not only speeds up the process but also produces more error free result. The aim is to use these techniques for determining the soil pH soil through image processing and regression. Different authors have already developed different methodologies for the predication of pH not only for soil but also for any kind solution. In 2011, a micro controller based digital pH meter has developed for human pH level by Mashud *et al.* (2011) where the output of the human pH is displayed with the help of seven segments display. To improve the performance of the device, authors have added an external oscillator circuit in their device. Again in 2011, fractal dimension method is implemented by Babu *et al*. (2011) using LabView to display physical properties of the soil. They have given the fractal dimension formula as follows

$$FD = \log (N(s)) / \log (1/s) \qquad (1)$$

For chemical characterization, 24-bit input color image is converted into 8-bit RGB image using LabVIEW. They have used pH index as follows

$$\text{Soil pH Index} = \text{Red} / \text{Green} / \text{Blue} \qquad (2)$$

From this, authors have calculated the pH index value of each pixel whose average gives the pH value of the sample. In this study, thresholding is done to convert the input images into binary form. Then Box counting method and the fractal dimension equation is used on the pixel value of the binary image for feature extraction. Later on Kumar *et al.* (2014) has used the same pH index as like Babu *et al.* (2011) for the calculation soil pH by extracting the red, green and blue signal of the soil images. In 2014, a fuzzy expert system has developed by Abu *et al.* (2014). They have control the level of soil pH of roses by altering pH of soil to an adequate level which helps to replace the addition of fertilizers directly and ensure healthy growth of the plants. The input for this fuzzy system is temperature, light intensity and humidity. The range for input variables are assigned as Temp: 0-1, Light Intensity: 0-5, Humidity: 0-1 and that of output variable is pH: 0-5. Again in 2014, an android application is developed by Ruiz *et al.* (2014) to measures the concentration of nitrite and also determines pH for paper micro fluidic devices. They have used a smartphone as a light source to capture the images with the help of Open CV library and extracted the Hue and Saturation of the images and related to pH and nitrite concentration. In 2016, soil pH is determined by Aziz *et al.* (2016) with the help neural network. In this study, they have calculated red, green and blue values of sample soil image and compared it with the other color values of the samples that are already stored in the database. They have found the minimum error to estimate the pH value of current sample using neural network.

MATERIALS AND METHODS

Methodology

Sample collection site: The soil images are collected by taking soil samples from the various areas of Guwahati under the supervision of department of civil engineering GIMT, Guwahati. The sample collection site lies in between 26.1445° N latitude and 91.7362° E longitude. The soil is taken from the topmost layer of the earth's surface with a depth of 6 inch in
. The collected soil samples are kept in white paper for image acquisition.

Image acquisition: The images (40) are captured with a Redmi 3s Prime with a resolution of 4160 x 2340 under visible light. Version: 6.0.1MMB29M, RAM: 3.00GB with F-stop=f/2, exposure time=1/60 sec, ISO speed=ISO 125, focal length=4 mm. To avoid the other light condition, we have captured the soil images from 10 inch ahead from the soil in a straight position as shown in the figure 3.

Soil pH calculation in laboratory: We have done our laboratory test in the Civil Engineering Department of GIMT Guwahati. The pH meter is calibrated using pH 7 buffer solution. Then, 20g of soil is weighed and transferred into 100ml beaker an again 40ml distilled water is added and stirred well with a glass rod.

This is allowed to stand for half an hour with intermittent stirring. To the soil water suspension in the beaker, the electrode is immersed and pH value is determined from the automatic display of the pH meter.

Processing of soil samples with image processing techniques

Image preprocessing: To improve the images which consist of undesired distortion, we have preprocessed the soil images with image enhancement techniques for contrast improvement, image resizing, image filtering.

Region of interest: The region of interest (ROI) is generated using the binary masking of the pixel points and extracting the required region.

Feature extraction: Every color of an image can be described by three vectors, one for every dimension in the color space. Here, we have used the RGB model, which is made up of red, green and blue intensity components, is the most common within digital imagery. The pixel values of each plane are extracted from the image. The RGB values are converted to HSV and then extracted.

We have calculated the HSV values for each soil image using the following formulas:

Mean H
Mean S
Mean V

RESULTS AND DISCUSSION

For result predication, we have proposed a color index which is based on the saturation values of the soil images as per the equation 3. Already Ruiz *et al.* (2014) has determined the pH of a paper micro fluidic devices using Hue and Saturation value of the images captured by the android device with the help of OpenCV library which permits to carry out in a more effective way the image processing needed for the detection of the sensing areas (Ruiz *et al.,* 2014) . But in our study, we have used only an android smartphone without any open source computer library so that it can be used by any rural farmers without any knowledge about computer library function.

Color index = Saturation   (HSV)   (3)

The following graph shows the value of color saturation index of 40 soil samples and the mean of the color saturation for the all samples is 0.044417255.
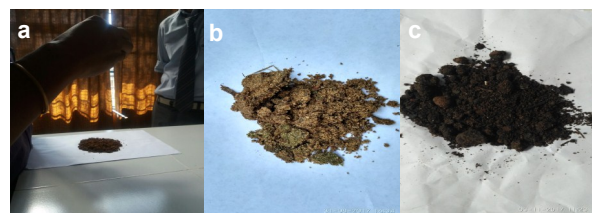
Initially we have performed the linear regression (LR) as suggested by Ruiz *et al.* (2014) to find the
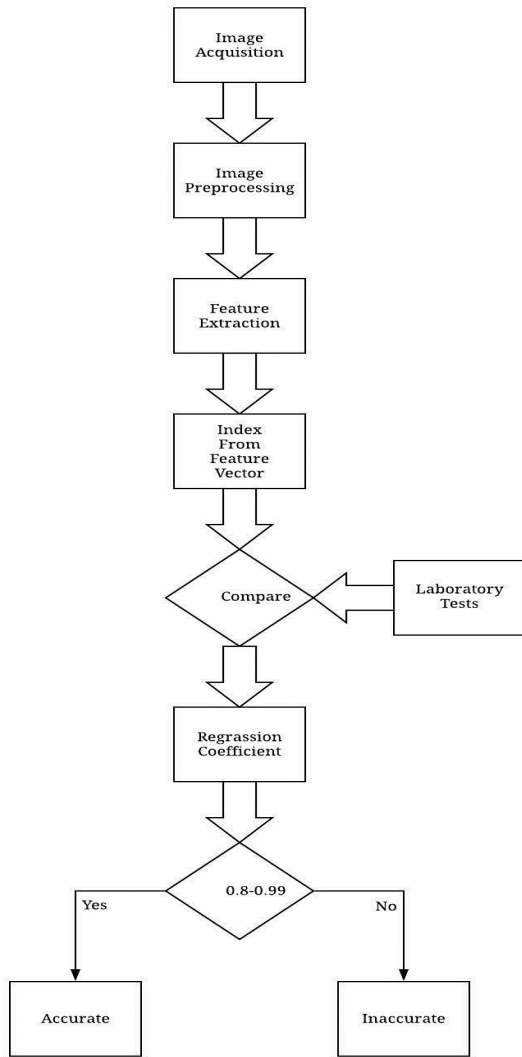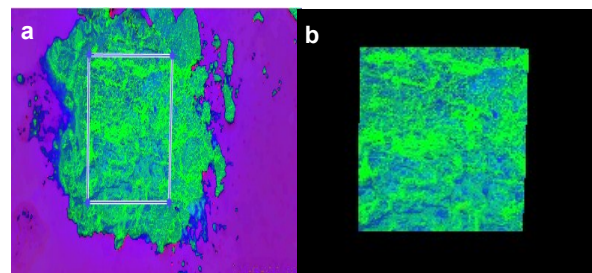

Fig. 1. *Flowchart of the proposed system.*

Fig. 2. *Area of Soil sample collection (source: Google Maps).*

Fig. 3. (a) *Soil images captured by Redmi 3S Prime smart phone. (b) Soil sample image (c) soil sample image.*

Fig. 4. (a) *ROI of the soil images (b) Result of ROI.*

correlation between soil pH values of laboratory test and color index of the sample images. We have found the regression equation for the given data as:

y = 46.01113X + 4.95556   (4)
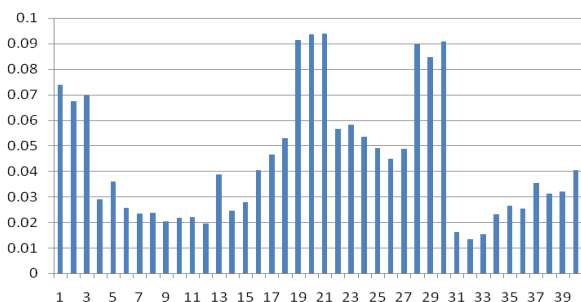
We have found the coefficient of regression analy-

sis as 0.8490795462 for soil pH but Ruiz *et al.* (2014) have found the coefficient of Fitting curves as 0.9984 for pH predication of paper micro fluidic device. Ruiz *et al.* (2014) determined the pH of the paper micro fluidic device after modified the sensing areas using phenol red and chlorophenol red reagents for uniform color as HSV coordinates take different values when the area is wet or dry. In our study, we have any not applied any reagents in soil for uniform color of the soil. We have collected the soil samples from the different region

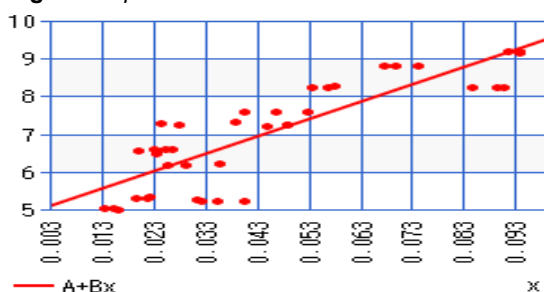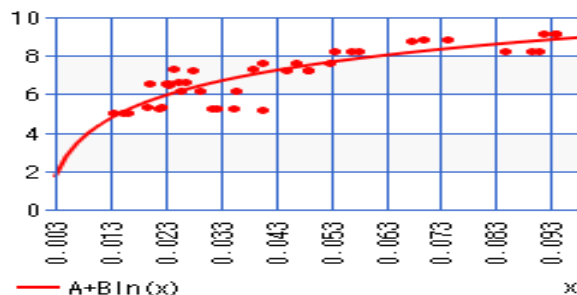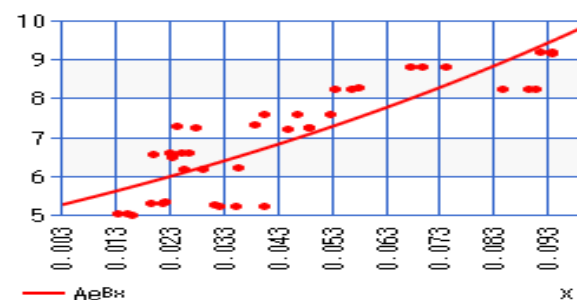**Fig. 5.** *Graphical values of color saturation.*

Fig. 6. *Linear Regression Graph for Saturation (X) vs. pH(Y) .*

Table 1. Region wise collected soil sample.

Sample	1-10	11-20	21-30	31-40
Region	Azara	Kahilipara	Silpukhuri	Narangi

Table 2. Values of pH for different soil samples.

Sample	pH	Sample	pH	Sample	pH	Sample	pH
1	8.82	11	5.33	21	9.19	31	5.01
2	8.80	12	5.32	22	8.25	32	5.05
3	8.83	13	7.33	23	8.27	33	5.02
4	6.20	14	7.30	24	8.24	34	6.60
5	6.22	15	7.27	25	7.26	35	6.62
6	6.20	16	7.60	26	7.22	36	6.61
7	6.50	17	7.60	27	7.27	37	5.23
8	6.54	18	7.61	28	8.26	38	5.25
9	6.55	19	9.18	29	8.25	39	5.24
10	5.30	20	9.16	30	8.26	40	5.21

Fig. 7. *Logarithmic regression.*

Fig. 8. *e-Exponential Regression .*

Fig. 9. *Quadratic regression.*

Table 3. Linear Regression calculation summary.

Sum of Values	Mean of Values	Sum of squares	Regression equation	Correlation Coefficient (r)
Sum of X =1.776 Sum of Y=279.97	Mean X =0.0444 Mean Y =6.9993	Sum of square=0.0241 Sum of Square =1.1095	y= b.X +a b = 1.11/0.02=46.01113 a =7-(46.01*0.04)=4.95556 y = 46.01113X + 4.95556	0.8490795462

Table 4. Regression summary of LR, ER and QR.

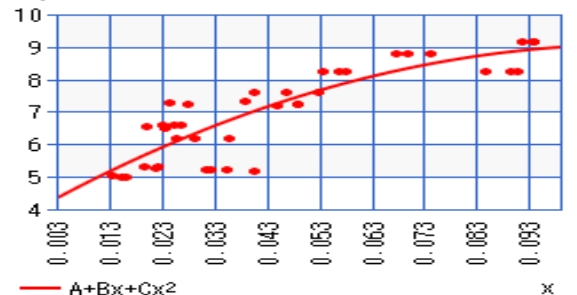| Function	Logarithmic regression (LR)	E exponential regression(ER)	Quadratic regression (QR)
Mean of x	0.0382146549	0.044417255	0.044417255
Mean of y	6.99925	6.869788531	6.99925
Correlation coefficient r	0.85703885	0.81872996	0.8629857624
A	13.7531505	5.14652615	4.08996968
B	2.0688698	6.50223366	88.099443
C	------	-------	-389.73594

of the guwahati so that a good mixture soils are comes out. Kumar *et al*. (2014) has concluded that soil pH values of dark brown colored soil are different to that of greenish and yellowish soil samples. The range of soil pH and pH index values are highest in that of the dark brown soils and lowest in yellowish soil samples. It means that their saturation values are more as compare to other types of soil that we have found in our study. They have found the range of soil pH and pH index values are 7.30-7.50 and 0.0070-0.0261, respectively in dark brown samples. Similarly, soil pH range varies from 6.80-7.04 and 5.58-6.58 in yellowish and greenish soil samples respectively. They have not found any statistical or machine learning based correlation in between the soil color index and pH values of the soil. All their assumption is in human perception level. Without any standard correlation, they found that RGB values in deep brown colored soil were 133-98-30 to 207-186-157 and its value in light yellowish soil 128-105-27 to 229-210-152 whereas in greenish soil RGB value ranged 152-122-52 to 189-164-113. Aziz *et al.* (2016) has used the soil dataset prepared by Kumar *et al.* (2014) in their experiment using neural network. They have found the coefficient of bets fit as $R^2 = 0.8$ which is low as compare to coefficient of linear regression of our study. Abu *et al*. (2014) has found their soil pH slightly different from the original values of soil pH but they have not explained the difference. They have found that soil with high temperature has low soil pH and vice versa for example soil pH is 2.16 when the temperature is 0.68. In our study also we have found less pH for those soils which are dry in nature as compare to wet soil due to the saturation of the soil image.

Finally, we have performed the logarithmic regression (LR), e exponential regression, (ER), quadratic regression (QR) with the following regression equation to increase the correlation in between for color saturation index with the value of laboratory ph value. In logarithmic regression (LR), we have found the correlation coefficient as 0.85703885 which is more as compared to linear regression result. But for the quadratic regression (QR), the value of coefficient is more as compare to all regression value.

y=A+Bln(x) (logarithmic regression)   (5)
y=AeBx (e-Exponential Regression)   (6)
y=A+Bx+Cx2 (Quadratic Regression)   (7)

Conclusion

Thus, after performing linear regression of the pH of the soil sample with the corresponding Saturation value, we obtained a correlation accuracy of 84.9% which concludes that the pH of the soil samples is linearly correlated with the saturation value. It is found that correlation coefficient R is more in case of quadratic regression which is 86%. Hence, quadratic regression gives a better result in case of soil pH in Guwahati area of Assam.

REFERENCES

Abu, M.A., Nasir, E.M.M. and Bala, C.R. (2014). Simulation of Soil PH Control system using Fuzzy Logic Method. *International Journal of Emerging Trends in Computer Image & Processing, 3(1):* 15-19

Aziz, M.M, Ahmed, D.R. and Abraham, B.F. (2016). Determine the pH of Soil by using Neural Network Based on Soil's Colour. *International Journal of Advanced Research in Computer science and Software Engineering,* 6(11):51-54

Babu, C.S.M. and Pandian, M.A. (2016). Determination of Chemical and Physical Characteristics of Soil using Digital Image Processing. *International Journal of Emerging Technology in Computer Science & Electronics,* 20(2):331-335

Kumar, V., Vimal, B.K., Kumar, R. and Kumar,M. (2014). Determination of soil pH by using digital image processing technique. *J. Appl. & Nat. Sci.,* 6(1):15-18

Masud, M.A., Mashud, M.A.A. and Islam, M.S. (2011). Design and development of microcontroller based digital PH meter. *Ulab Journal of Science & Engineering,* 2(2):31-35

Ruiz, N. L., Curto, V.F., Erenas, M. M., Lopez, F. B., Diamond, D., Lopez, A. J. P. and Valley. A.F.C (2014). Smartphone-Based Simultaneous pH and Nitrite Colorimetric Determination for Paper Microfluidic Devices. *Analytical Chemistry*, 86(19):1-23