Application of Colorimetry to determine Soil Fertility through Naive Bayes Classification Algorithm

Surili Agarwal, Neha Bhangale, Kameya Dhanure, Shreeya Gavhane, V.A.Chakkarwar,
Dr. M.B.Nagori
Dept. of Computer Science and Engineering
Government College of Engineering, Aurangabad
Maharashtra, India
agarwalsurili2203@gmail.com, bhangaleneha9092@gmail.com, ind.kameya@gmail.com,
shreeyagavhane97@gmail.com, vrush.a143@gmail.com, meghana@geca.ac.in

***Abstract*— Fertility of the soil is considered most important criterion in any agriculture practice. Nutrients present in the soil define its fertility. Mineral nutrients such as Nitrogen (N), Potassium (K), Phosphorous (P) are vital for plant growth and food production. Lack of adequate knowledge amongst the farmers about the various parameters in farming like the soil fertility, amount of fertilizer to be used, leads to degradation of the overall soil quality. In this paper, we have represented a system to test the soil fertility by using the principal of colorimetry. Colorimetry is a technique in which we measure the amount of light absorbed by the color developed in the sample. An aqueous solution of the soil sample is prepared using extracting agents and is subjected to the photodiodes of the color sensor. The solution develops a color due to reaction of nutrients in the soil with chemicals. The output by the color sensor is calibrated with standard values present in the database. To verify the results obtained by the color sensor we use the Naive Bayes classification algorithm. This algorithm classifies the intensity values of the soil solutions into three class labels namely low, medium, high. After applying the Naive Bayes classifier, we can predict the accuracy of the intended system. The intended system is thus beneficial to reduce the time required for testing the soil fertility and determining the accuracy of our results.**

Keywords*— *Colorimetry, Naïve Bayes, Nutrients, Soil fertility

I. INTRODUCTION

CONTRIBUTION of agriculture sector in Indian economy is quite prominent and is much greater than world's average. Nearly 18.1% of the Indian annual gross domestic product is contributed by Agriculture[1]. However Farming in India is still done by adopting traditional practices which generally makes use of approximations. The growth and productivity of crops has been stagnant in recent years due to degradation in quality of the environmental consequences in the form of depleting water table and emission of greenhouse gases. With the emergence of technology, there has been a significant increase in monitoring various agricultural parameters like soil moisture, humidity, temperature etc. [2], in order to improve the farm field. The fertility of soil is a crucial factor in increasing the crop production. Soil fertility simply means the amount of nutrients content existing in the soil. There are 16 elements in the soil that are considered essential for any plant growth. These nutrients can be categorized into two classes namely macronutrients and micronutrients. Macronutrients are are generally required in large quantity in the soil. Nitrogen (N), Phosphorous (P), Potassium (K) are some of the major macronutrients. Micronutrients on the other hand are required in relatively small amounts. Calcium, magnesium, zinc, copper, iron, boron etc. are examples of micronutrients.

The quantity of macronutrients defines the soil fertility. Different crops require different quantity of macronutrients [3]. Depending upon the deficiency of a particular nutrient for a particular crop in a given region the amount of fertilizer is determined. Thus the quantity of fertilizer to be used depends upon the soil fertility [3], crop to be grown and type of soil. In most conventional practices the farmer estimate the amount of fertilizer to be used without actually testing the soil. Research shows that any field should be retested before cultivation, at least every four years to prevent the nutrient deficiencies and maintain the nutrient balance.

The routine soil testing program in laboratories take a considerable amount of time to generate the results [4]. During the period from testing to getting results, the overall soil fertility changes as it is often subjected to changing weather conditions. Time becomes a crucial factor in case for determination of soil nutrients.

This system works on the principle of colorimetry. An aqueous solution of the soil sample with certain chemicals is prepared and is subjected to the color sensor. A color is developed due to reaction of nutrients in the soil with chemicals. These chemicals are called extracting agents. The output by the color sensor is compared with the standard results stored in a database. After comparison the deficiency is displayed. To verify the correctness of the system, the Naïve Bayes classification algorithm is used. The key feature of our system is to reduce the time required for testing the soil fertility and determining the accuracy of our results.

II. RELATED WORK

In a study on Electrochemical Sensors to determine Soil fertility [5], we found that the Electrochemical Sensors consist of Ion Selective Electrode (ISE) and Ion Selective Field Transistor(ISFET). These electrodes are used to measure the concentration of ion in the solution. Electrochemical sensors can be integrated onto one chip as a sensor array to provide a feasible approach of multi-targets simultaneous detection. The main challenge faced here is reliability of sensor array, that is, to avoid or diminish the interferences from other ions while using a sensor array for simultaneous detection.

Another technique to determine soil fertility using optical method is the spectroscopy [6]. Laser Induced Florescence

Spectroscopy (LIFS) in which the solution absorbs radiations usually from the UV range. This method is reliable but complex. Also the number of samples that can be tested by the system are limited. The system is very specific as the range of wavelength is only in the UV band. Another spectroscopy method is Near Infrared Spectroscopy (NIR), which is based on absorption of the electromagnetic radiations. The limitation of this method is that it can only detect wavelengths in the range 780-2500nm. NIR is costly, time consuming and complex.

There is another existing technique to measure soil nutrients using fiber optic sensor [3] which works on colorimetric principle. In this technique, the sensor is built using multimode, plastic optical fibers. The sensor consists of seven fibers arranged in circular configuration with one fiber at the center. The central fiber acts as receiving fiber and surrounding six fibers acts as transmitting fibers. With the help of multimode optical fibers light is incident on the aqueous solution of given soil sample. The reflected light is captured by the receiving fiber of the sensor and converted to electrical signal. Then the electrical output is calibrated in terms of the status of nutrients present in the soil. The techniques that are using fiber optic sensors are reliable but costly and time-consuming.

As compared to these methods, the intended system methodology is relatively less costly, less complex and the time required to produce results is also less.

III. BLOCK DIAGRAM

Fig.1 depicts the block diagram of the intended system. The soil nutrient detection system consists of Arduino micro controller, color sensor, regulated power supply, a computer system and soil sample.

As shown in Fig. 1, The Arduino UNO board is connected to the computer system via cable provided with Arduino Uno board. The results can be directly viewed on the computer screen through the Arduino software installed on computer. The color sensor is connected to the Arduino board using jumper wires.

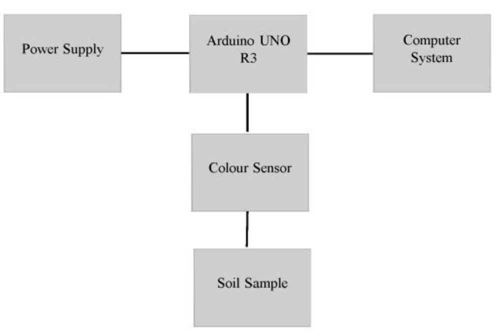

Fig. 1 Block Diagram

A. Color sensor

Fig. 2 TCS 3200 color sensor

Color sensor TCS3200 is a hardware device which senses color light using an 8*8 array of silicon photodiodes [6] with red, green and blue filters. These values from the array are then transformed to the square frequency which is directly proportional to the intensity of light. The Arduino micro controller can read the output of the sensor and generate results for the color in the form of intensity values.

B. Arduino Uno R3

Fig. 3 Arduino Uno R3 board

Arduino UNO R3 is an open source [7] platform that designs micro controller based projects that can sense and control objects in physical world. Arduino consists of both a physical circuit board and a software or IDE. This IDE can be utilized for writing as well as uploading the code to the physical circuit board.

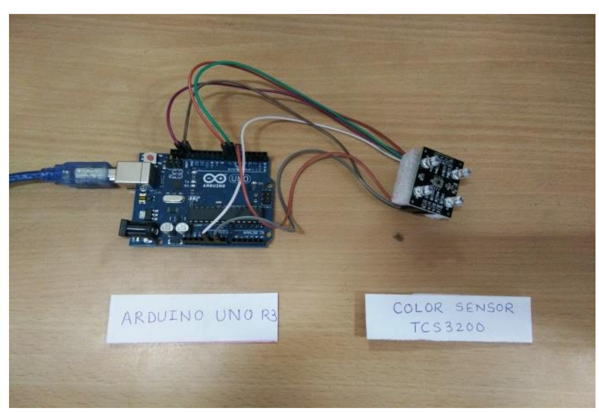

Fig. 3 Circuit connections

IV. Proposed System Methodology

Fig. 4 represents the steps carried out to determine the soil fertility. The steps can be broken into four major components namely soil sampling, soil pretreatment, actual measurement of NPK using color sensors and verification using Naïve Bayes classification algorithm.

A. Soil Sampling

Selection of soil samples is most crucial step for the detection of N P K as well as for fertilizer predictions [8]. Hence selection should be done by considering the analytical cost, equipments available, time required and also field fertilization history. Soil sampling is done in process format .At first stage field area is chosen from where the soil is to be tested and samples are to be collected. The area from where the soil samples are to be collected should possess same fertility and crop history. Also the characteristics such as color, texture, slope of the soil should be similar. At the second stage the tool required for sampling comes into picture i.e. a stainless steel probe is most significant tool being used. Vehicle mounted hydraulic probe is also a better choice under adverse conditions. Different tools which include spade, shovel, sample bags, plastic (non-vulnerable) bucket and markers for identifying samples on sample bags are used. By using tools, sampling depth is estimated at third stage. The soil from (surface to six inches) is selected for testing which has the most fertilizer applications.

Before digging up the soil the organic matter (weeds, stones, hay etc.) on the surface of the soil is removed successfully. To detect nitrogen in the soil, a deeper range i.e. from (6 inch to 24 inch) of soil in V format is collected. To result out the proper analysis, the soil samples are collected in "Zigzag" [9] manner so that there would be collection of samples of soil from different sections of field. The amount of sample to be collected comprises of fourth and most profound stage. Approximately (15-20) samples are collected from minimum possible depth of 24 inches. The core of soil is divided into two portions, first portion (0 inch -6 inch) and second portion (6 inch - 24 inch). The samples of first portion is collected in bucket and mixed well and hence the surface samples are assembled in the plastic bags .Similarly the second portion i.e the subsurface samples (6 inch - 24 inch) are gathered is similar way. If the field land is variable then additional samples to a (depth > 24 inch to 48 inch) is endorsed .Before pouring the samples in plastic bags it is well dried by placing the whole sample on paper and letting it dry on room temperature.

Hence now the soil samples are organized for further procedure.

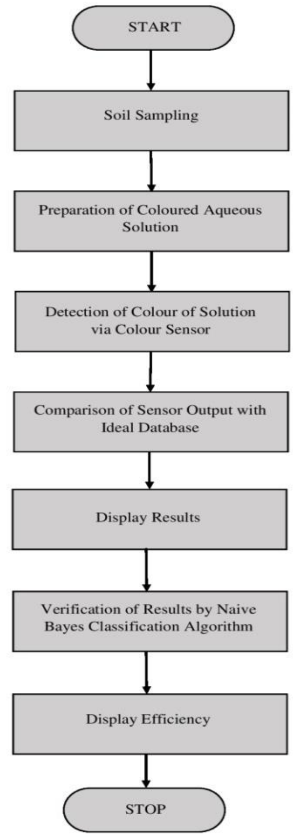

Fig. 4 Working Flow Diagram

B. Soil Pretreatment

This step includes preparation of the colored aqueous solution. We are using a Soil Testing kit to check the fertility of a particular soil sample. After collecting soil sample, add approximately 1gm of the sample to test tube A with the help of funnel. The second step is to add 6 ml of water in the test tube A. Put the cap on the tube and shake it for a minute. We have to keep the test tube A still for 5 minutes so that soil will settle down and we get a clear solution on top of the test tube. The next step is to take 2 ml of clear solution from test tube A to another test tube (test tube B) using a dropper. Make sure that there will not be any soil traces in the test tube B. Now add one nitrogen capsule from the soil testing kit. The nitrogen capsule usually contains KCl as an extracting agent. Put the cap on the tube B and shake it gently. We have to keep the test tube B still for 5 minutes to allow the solution to develop a color. The color sensor now senses the color intensity existing in the solution of test tube B and gives results in the form of three values which correspond to red, green and blue. These values are then analyzed, displayed and categorized into three class labels namely low, medium and high. We are using the same procedure to test phosphorous and potassium in a soil sample.

Measurement of NPK Using Color Sensor

For each sample to be tested, the solution is prepared by the manual process as stated above. TCS3200 programmable color light to frequency convertor that combines configurable Silicon photodiodes and current to frequency convertor on a single monolithic CMOS integrated circuit. The color Sensor TCS3200 can detect and measure all colors of the visible spectrum. The output generated by the color sensor is in the form of a square wave. The frequency of this square wave is directly proportional to intensity of the light. The color sensor consists of two main components the TCS3200 RGB sensor chip and 4 white LED's. The light emitting from the LED's is incident on the aqueous solution of the soil sample. The soil sample reflects the light of certain wavelength which is detected by the array of photo detectors in the color sensor. The sensor output is calibrated with standard values present in the database. The deficiency of the major nutrient can be easily measured because of the color sensor output displayed on the screen as low/ medium/ high.

C. Verification by Naïve Bayes Classification Algorithm

Naïve Bayes is a classification algorithm that employs the Bayesian algorithm to perform classification [10]. It presumes that the occurrence of a certain attribute is independent of the occurrence of other attributes. That is why this algorithm is called as "Naïve". It is a widely used approach for classification because it gives reasonable performance. It allots the class labels to the problem instance employing the concept of conditional probability.

$$P(C|F) = \frac{P(C).\ P(F|C)}{P(D)}$$

This formula computes the probability of occurrence of event C such that event F has already occurred [11]. It is also called as Bayes theorem. In a classification problem, there are classes and features. C1, C2, ……Ck being the classes, Naive Bayes classification algorithm intends to compute the conditional probability that an item with feature vector y1, y2,…..yn belongs to class Ci.

$$P(Ci|y1, y2 \ldots\ldots yn) = \frac{P(y1, y2 \ldots\ldots yn \mid Ci)\ .\ P(Ci)}{P(y1, y2, \ldots\ldots yn)}$$

This formula computes the probability of the item with feature vector <y1, y2……yn> belonging to class Ci. This probability is calculated for all the classes ranging from 1 to k. The class for which the value of probability is highest is the one to which the item with the given features belongs. Thus, it computes the probabilities that the item with given features belongs to particular class. Based on the computed probability values, it allots the class label to the problem instance.

We have measured the status (low/ medium/ high) of nutrients in the soil using color sensor. Now we have to verify whether the results are accurate or not. So we use Naïve Bayes classifier for this purpose. The output data of the color sensor will act as the test data for Naïve Bayes algorithm. In the test data we specify the list of samples that were tested using color sensor that is intensity values of the samples for a particular nutrient and the corresponding status (low/ medium/ high) of the nutrient content in soil. Thus here Low, Medium, High are class labels and intensity belongs to feature vector. The ideal color intensity values for Low, Medium, High for a particular nutrient are used and the classifier is trained based on some part of data. The Naïve Bayes classifier will determine how many samples were correctly classified i.e how many samples actually belong to correct class and how many samples were wrongly classified. It will also display the percentage of correctly classified instances which states the accuracy of the system.

V. Experimental Results

TABLE I. Threshold color intensity values. [3]

Component	**Low**	**Medium**	**High**
Nitrogen	$x < 15$	$15 < x \leq 20$	$20 < x \leq 25$
Phosphorous	$16 < x \leq 20$	$20 < x \leq 35$	$35 < x \leq 50$
Potassium	$20 < x \leq 25$	$25 < x \leq 40$	$50 < x \leq 60$

TABLE II. Training Data

Sample No.	N (Color Intensity)	P (Color Intensity)	K (Color Intensity)	Status (N, P, K)
1	12	30	59	(low, medium, high)
2	10	17	22	(low, low, low)
3	18	18	55	(medium, low, high)
4	22	33	23	(high, medium, low)
5	19	31	57	(medium, medium, high)
6	25	40	35	(high, high, medium)
7	19	46	37	(medium, high, medium)
8	24	48	58	(high, high, high)
9	18	18	38	(medium, low, medium)
10	10	39	23	(low, high, low)

TABLE III. Test data.

Sample No.	N (Color Intensity)	P (Color Intensity)	K (Color Intensity)	Ideal Status (N, P, K)	Classified results
1	22	23	59	(high, medium, high)	(low, medium, high)
2	12	17	53	(low, low, high)	(low, high, high)
3	10	29	52	(low, medium, high)	(low, medium, low)
4	17	17	54	(medium, low, high)	(medium, low, high)
5	22	37	22	(high, high, low)	(high, high, low)
6	19	43	33	(medium, high, medium)	(medium, low, medium)
7	16	17	22	(medium, low, low)	(medium, low, low)
8	9	17	25	(low, low, low)	(low, low, low)
9	31	46	53	(medium, high, high)	(medium, high, high)
10	19	32	25	(medium, medium, low)	(medium, medium, low)

The threshold color intensity values for N, P, K contents in the soil are depicted in TABLE I. It provides the range of color intensity values that is low, medium, high for N, P, K.

The dataset is split into training data and test data. The subset of training data and test data are depicted in TABLE II and TABLE III respectively. Using Naïve Bayes classifier, we assess how accurate the system is in classifying the nutrient status. A dataset of 100 tuples is divided in the ratio 3:2 as training and test data respectively. Out of the 40 instances of test data, 32 instances were correctly classified. That is if the nutrient status is low then it is classified as low and likewise. From this, the accuracy of the system is determined. Accuracy is simply the ratio of correctly classified instances to the total number of observations.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}}$$

Where,

TP (True Positive) means the positive tuples that are truly labeled by the classifier.

TN (True Negative) means the negative tuples that are truly labeled by the classifier.

P + N refers to the total number of observations.[12]

For our model, we have got the accuracy of 0.80 which means our model is 80% accurate.

VI. Conclusion

The proposed system detects the soil fertility by using the principle of colorimetry and classification algorithm. The system will help farmers determine the Soil fertility with accuracy and reduces the time involved in the conventional soil testing Methodology. When compared with other soil testing methods using colorimetry like fiber optics sensor and spectroscopy, the proposed system is more reliable and affordable. As we are using Naive Bayes Classification algorithm, the accuracy of the system in determining the status of major nutrient contents in the soil is verified. The system was tested using soil samples with confining soil types typically black soils with clayey texture and dark brown soil with blocky structure, clayey texture. The texture, color, structure of the soil samples were restricted to few category. Also the dataset used for the analysis of the proposed methodology was limited to the spatial soil characteristics of the region .For large soil datasets comprising of entire country, efficient methods could be created that utilize Data Mining/Big Data techniques to enhance the exactness of classification. Also the system does not provide exact quantification of N, P, K contents in the soil.

VII. Future Scope

Our analyzed and implemented system is liable to detect the NPK, the most important macronutrients in the soil. But it can be further evolved into better version by adding the fertilizer prediction required for the particular soil type in a specific region. Quantification of exact values of NPK along with fertilizer recommendation can be the better future implementation of the proposed system.

Acknowledgment

Authors are grateful to the project research guide Prof. V.A.Chakkarwar for inspiring us on the context and motivating us to implement our research in a proper and courteous way. We would also like to thank Meghana.B. Nagori for helping us to realize the realistic approach with respect to our research work and paved in organizing the structure of the research paper. We would also like to thank our Head of Department Prof. Vivek.P.Kshirsagar for conducting timely reviews and surveillance which helped in directing our ways out. Our parents for providing the moral support and developing patience in our minds. The last but not the least our almighty God for letting us into the world and providing us an opportunity to work under our magnanimous research guide.

References

[1] Amrutha A, Lekha R, A Sreedevi, "Automatic Soil Nutrient Detection and Fertilizer Dispensary System," *International Conference on Robotics: Current Trends and Future Challenges (RCTFC)*, 2016.

[2] Mr. Gaikwad S.V, Prof. Galande S.G, "Measurement of NPK, Temperature, Moisture, Humidity using WSN," *International Journal of Engineering Research and Applications*, vol. 5, pp. 84-89, August 2015.

[3] Deepa V. Ramane, Supriya S. Patil, "Detection of NPK nutrients of soil using Fibre Optic Sensor," *International Journal of Research in Advent*

Technology Special Issue, National Conference "ACGT 2015", 13-14 February 2015.

[4] Methods Manual – "Soil Testing in India", Dept. of Agriculture & Coop, Ministry of Agriculture, Govt of India, New Delhi, Jan. 2011.

[5] Jianhan Lin, Maohua Wang, Miao Zhang, Yane Zhang, Li Chen, "Electrochemical sensors for soil nutrient detection: Opportunity and Challenge," Key Laboratory of Modern Precision Agriculture System Integration, Ministry of Education, Beijing, China, 100083

[6] Kunhimohammed C. K, Muhammed Saifudeen K. K, Sahna S, Gokul M. S, Shaeez Usman Abdulla, "Automatic Color Sorting Machine Using TCS320 Color Sensor and PIC Microcontroller," *International Journal of Research and Innovations in Science and Technology*, vol. 2, 2015.

[7] Salve Akshay B, Sagar Sonali S, Patne Mahesh L, Jangam Omkar R, "Soil Nutrient Identification Using Arduino and Electrochemical Sensor," *International Research Journal of Engineering and Technology*, vol. 05, February 2018.

[8] Dharesh Vadalia, Minal Vaity, Krutika Tawate, Dynaneshwar Kapse, " Real Time soil fertility analyzer and crop prediction," *International Research Journal of Engineering and Technology*, vol. 04, 2017.

[9] Carrie A.M. Laboski, John B. Peters, *Nutrient application guidelines for field, vegetable and fruit crops in Wisconsin*.

[10] V. Rajeswari , K. Arunesh, "Analyzing Soil Data using Data Mining Classification Techniques," *Indian Journal of Science and Technology*, vol. 9, May 2016.

[11] Preety, Sunny Dahiya, "Sentiment Analysis Using SVM and Naïve Bayes Algorithm" *International Journal of Computer Science and Mobile Computing*, vol. 4, pp. 212-219, September 2015.

[12] Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining Concepts and Techniques*