IoT-based Soil Nutrient Analyser using Gaussian Process Regression

Isuru Senevirathne
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
senevirathnegit.21@uom.lk

Thanuja Ambegoda
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
thanuja@cse.mrt.ac.lk

Ruchira Wijesena
Sri Lanka Institute of Nanotechnology (SLINTEC)
Homagama, Sri Lanka
ruchiraw@slintec.lk

Indika Perera
Department of Computer Science and Engineering
University of Moratuwa
Moratuwa, Sri Lanka
indika@cse.mrt.ac.lk

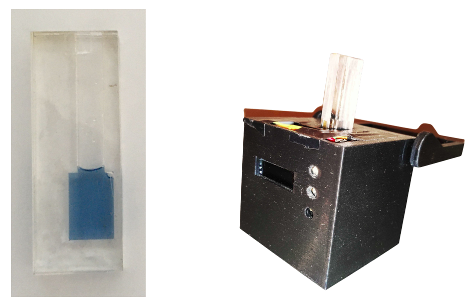
2022 2nd International Conference on Advanced Research in Computing (ICARC) | 978-1-6654-0741-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICARC54489.2022.9754105

***Abstract*—Delivering the right amount of fertilizer with the optimal nutrient compositions (N, P, K) is critical for high yields and to sustain long-term soil quality. Soil condition and the fertilizer management in agricultural fields has been a challenge due to the lack of cost effective soil nutrient sensing methods that can rapidly give the results. We have produced an inexpensive and portable IoT-based soil nutrient analyzer that estimates the level of soil nutrient contents using colorimetry and machine learning. The proposed device estimates the nutrient content by initially detecting the color of a microfluidic cassette which contains chemical reagents and the soil sample. The use of microfluidic cassettes drastically reduces the required amount of chemical reagents, thereby brings down the cost of a test. This color readings that are determined by the nutrient concentrations present in the soil sample are fed into a supervised machine learning model based on Gaussian Process Regression (GPR), that runs on a mobile app which predicts the nutrient concentrations of the soil sample. The GPR model also provides the variance of the estimates. In this paper, we present the nutrient sensing IoT device and the nutrient prediction model based on the Gaussian process regression. We validate our device and the prediction model with experimental data where we achieve 93% accuracy in predicting the soil nutrient concentration level (Low, Medium, High).**

***Index Terms*—Soil nutrient analysis, fertilizer recommendation, smart agriculture, edge AI**

I. Introduction

Soil quality in agricultural fields has become immensely unpredictable, mismanaged and below the required standards due to the haphazard application of fertilizer to fields without a proper assessment of the prevailing soil condition. Most of the farmers apply fertilizer by guess work, experience, and an unorganized knowledge system which they inherit. The main intention of this research work is to provide a reliable solution to analyse the main nutrient components of the soil in a short time with a low cost. The proposed soil nutrient analyser provides the current nutrient concentrations in the tested field, which is useful for the agrarians and the farmers to derive recommendations for fertilizer input required for that specific field, to maximize the crop yield and for the long term sustainability of soil quality.



(a) (b)

Fig. 1. The proposed device to measure the soil nutrient concentrations. (a) Cassette which contains a 300 µL sample of the colored solution which is being read by the color sensor device. (b) The device and insertion of the cassette to the device.

The proposed method is able to measure the Nitrogen, Phosphorus and Potassium nutrient contents (NPK) in three separate trials. These are the major nutrient requirements when evaluating the soil condition. The output given by the nutrient analyzer would be rating shown as Low, Medium or High in the relevant nutrient. This reading can be used by the agrarians and farmers to arrive at the fertilizer recommendation. The color reading process of the developed color in the special cassette provided with the device takes less than one minute.

To estimate the nutrient contents, a chemical procedure is followed by the color detection of the developed solution, which then the color values are fed into a supervised learning model based on non-parametric Gaussian process regression.

Our contributions are as follows:

- A novel cost effective IoT and ML based solution for on site soil nutrient analysis based on a colorimetric device that uses absorption properties. The prototype

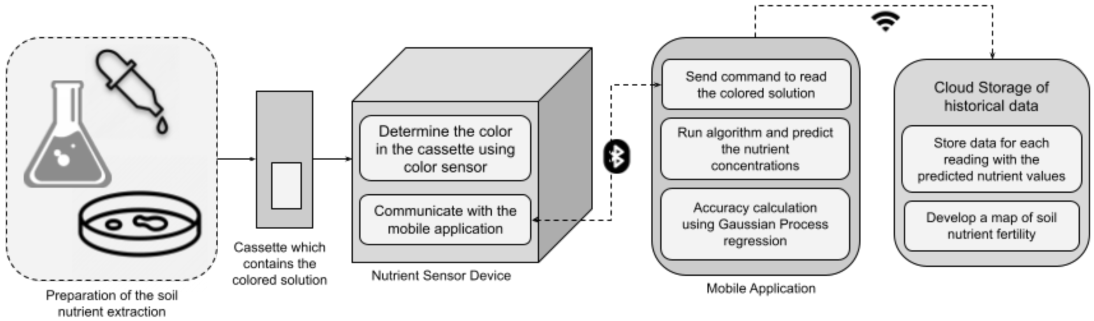


Fig. 2. System Architecture: The main components of the proposed system consists of nutrient sensor device, mobile application and the cloud storage. Soil measurement is carried out after preparing the soil sample and inserting it in the cassette to be measured from the nutrient sensor device.

costs around USD 30 per unit, which could be brought down to around USD 15 per unit when mass produced.

- Reliable colorimetry based nutrient readouts and their variance estimates using Gaussian Process Regression, which we have validated experimentally. We have obtained a reliable variance estimate of 0.34. GPR allows the model to be calibrated for different types of color readouts only using a few data points.
- A smartphone application that connects to the colorimetry device wirelessly, for easy calibration, nutrient content readouts and cloud services
- Validation of our prototype using a set of solutions with known phosphorous content

We have produced a prototype that is able to detect the color of the solution inside the cassette and evaluate the nutrient concentration related to the detected color and its intensity. The preparation of chemical reagents is not in the scope of the work presented in this paper. Although in this paper we have only validated our setup for measuring phosphorous, it generalizes to any colorimetry-based measurement that can produce a color corresponding to the nutrient we are interested in measuring.

II. Related work

There have been several recent approaches where colorimetry was used to determine the soil nutrient contents. In [1], reflection-based colorimetry is carried out using TCS4371 color sensor and LDR. The obtained values are fed into a VB.NET application which suggests fertilizer amounts to be added. Soil test kit based testing procedures for each NPK and pH readings and accuracy was validated by the human readings. Results convey that they computed no errors in the device and human readings when using Chi-squared.

In [2], the quantification of the soil nutrient levels using colorimetry and image processing techniques was done by developing soil test kits and rapid soil testers. A web camera and light controlled box is used to capture the color of the solutions inside the test tubes. The soil test kits obtain color readings for different nutrients such as Nitrogen, Phosphorus, Potassium, Zinc, Calcium, and Magnesium. They have developed and trained an artificial neural network to identify the colors in the solutions and hence predict the nutrient concentrations. This approach needs image noise reduction, brightness adjustment and segmentation of the region of interest after capturing the test tube image. Their performance shows 100% accuracy for the obtained data. Authors have recommended further improvements to increase the test sample size to yield more accurate results.

Another system was proposed in [3], where soil moisture is measured using a soil moisture sensor(FC28), and NPK values are measured using colorimetry. In this method, they use reflection based color sensing where they have used a TCS3200 color sensor to detect the red, green and blue colors using the filters available in the sensor. Resistive type moisture sensor is used and an analog value (0 to 1023) is obtained and converted to a percentage. Sensor data is sent to Firebase realtime database and the mobile application displays the sensed values. The application predicts the NPK values in kg/ha and then classifies them to Low, Medium, High levels.

Similar work was done in [4] , using the reflection based solution color sensing with TCS3200 color sensor. The nutrient levels are determined by the color intensity of the developed chemical solution using the reflection based sensing. The nutrient readings are classified to Low, Medium and High. Naive Bayesian Classification is used to calculate the accuracy of the system. They have achieved 80% accuracy in detecting the nutrient concentration as per the paper.

There more work done in color sensing, where in [5], analysis of fruits by their color, colorimetry principles and applications have been used. This work also has used a TCS34725 color sensor, however the color space which has been used is the HSV color space instead of the RGB color space. It is stated that HSV color space is more close to the human vision rather than RGB color space.

Most of the relevant work consists of colorimetry techniques to measure the chemical solution colors, however since these chemical solutions are in liquid state, absorption based sensing mechanism is more accurate compared to reflection based sensing mechanism which is currently used. Furthermore, the device portability and onsite reading is challenging in most of these work.

III. METHODOLOGY

We have proposed a system which can read the nutrient levels using the color measurements of the chemically processed soil sample. Our approach is to get the color readings from the color sensor in the device and then transmit it to the mobile application, where the pre-trained Gaussian process regression model predicts the nutrient concentration in the soil sample. Accurate measurement of the color in a solution is a challenging task, hence a colorimetric method was used based on light absorption.

A. Soil test kit

To perform colorimetric quantification of soil nutrients, we have to first used a set of chemical reagents in predefined quantities to react with a dissolved soil sample. The exact composition of the chemical reagents would depend on the nutrient for which the test is performed. The set of chemical reagents specified and the apparatus (microfluidic cassette) in which the reaction is taking place will be here after referred to as the *soil test kit*.

This step will cause a colorization that would be quantified by the device we propose, which would estimate the nutrient content reliably. Our device is independent of the choice of the chemical reagents because it can be calibrated for different types of reagents as long as a unique colorization is caused by the presence of the nutrients we are interested in.

The soil test kits contain soil extraction capability for main nutrients such as nitrogen, phosphorus and potassium. As the output of this soil nutrient extraction kit, a color area is developed in the provided transparent cassette. The color and its intensity is used to determine the nutrient concentration of the provided soil sample, which is currently done comparing the color charts with the human eye. Digitizing the color determination and the soil nutrient prediction is the main focus area in this work which provides the base to build a digitized system to manage the soil nutrient levels in a more scalable and consistent manner.

In the system level view Figure 2, the main component is the soil nutrient sensor device which contains a color sensor and Bluetooth enabled microcontroller circuit. The developed color is measured through this color sensor using the absorption based sensing mechanism. There are several methods to determine the color of solids such as reflection based sensing, though it is challenging to determine a liquid color accurately in this method. Absorption based sensing mechanism based on the Beer-Lambert law is developed to accurately measure the developed color in the cassette. The color readings are then transmitted to the mobile application, where a machine learning model is used to predict the nutrient concentrations based on the derived color readings of the developed color in the cassette.

The relationship between the color readings and the nutrient concentrations are estimated using a supervised learning method known as Gaussian process regression. A suitable kernel method is used such that it can accurately predict the relationship. Gaussian process can predict with a variance estimation which can elaborate the confidence level in the prediction. Gaussian process regression is particularly useful to model the limited training data and samples.

B. Colorimetric nutrient reader

1) Hardware design and operating principle: Colorimetry is the technique which is used to measure the concentrations of the colored compounds in solution using the Beer-Lambert law. In spectrometry, there are two main methods for measuring the color which are reflection based method and the absorption based method. When considering the behaviour of the chemical solutions as shown in Figure 3, when the solution is placed in a transparent cassette, the amount of light reflected from the cassette is much less compared to the light transmitted through the cassette. Therefore we chose the absorption based color sensing method over the reflection based sensing method.

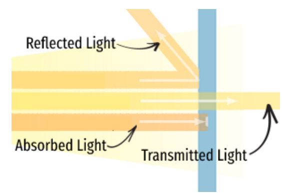
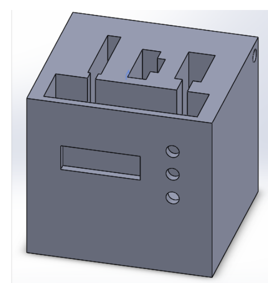


Fig. 3. Transmission patterns of light through a solution. Reflectance is very low compared to the transmittance and absorption in solutions.

(a) Prototype device which fully blocks the ambient light



(b) 3D model of the prototype device

Fig. 4. Soil nutrient sensor device enclosure which can block the ambient light

A fully enclosed enclosure as in Figure 4 which can effectively block all the ambient lights and focus the RGB LED light towards cassette was designed. This light is transmitted towards the color sensor which is used for detecting the color and its intensity. The color sensor used is TCS34725 [6] light to digital converter which has an IR blocking filter. The color sensor consists of a 3X4 photo diode array, composed of red-filtered, green-filtered, blue-filtered and unfiltered photo diodes to detect the colors accurately.

To improve the individual color detection in RGB color space, the light beam was designed using an array of RGB LEDs such that three red, green, blue colors can be generated separately. Thus, for a single trial, three color readings were obtained using the three light beams. Therefore, three color

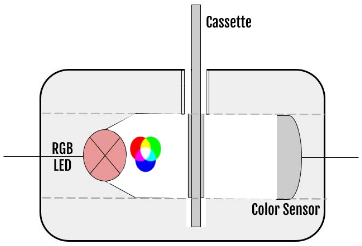


Fig. 5. LED, Cassette and Color sensor setup

reading values are obtained for each reading. The sensor and LED setup which can detect the absorption and transmission light is shown in Figure 5.

The main controller of the device is ESP32-Wroom-32 microcontroller which has powerful features such as two 32-bit microprocessors, I2C/SPI/UART data wired communication interfaces, as well as wireless communication interfaces such as Bluetooth and WiFi. Color sensor is connected to ESP32 using the I2̂C interface. The Bluetooth serial is used to connect with the mobile application virtually. To make the device portable, a Li-ion battery is used to power the device which is regulated to 3.3V with $1000\,\mathrm{mAh}$ and charged using the TP4056 battery charging circuit as shown in the device block diagram in Figure 6.

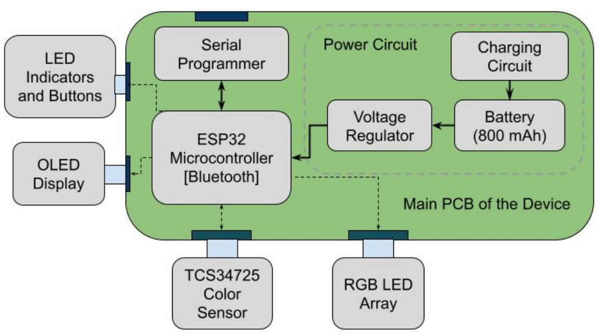


Fig. 6. Device block diagram

2) Data Collection: For the initial data collection, the Transchem soil test kit [7] was used which is shown in Figure 7. Test samples were created by making specific concentrations that were prepared using laboratory available chemicals equivalent to soil nutrients mixing with distilled water. The reagents are added as required to develop the colored solutions for each nutrient measurement. The required reagent amounts are calculated and measured from the digital pipettes and added to the prepared chemical samples. The prepared samples are kept for 10-20 minutes to complete the color development before the color measurement using the device.

The prepared colored solutions are added to the testing cassette using the digital pipette by measuring the $300\,\mu\mathrm{L}$ which is required to fill the color measuring area.

Amount of Nitrogen available in the soil is to be determined by the Nitrate ($NO_3{}^-$) or Nitrite ($NO_2{}^-$) in soil. Using cadmium reduction procedure, which would produce N-l-naphthyl-ethylenediarninedihydrochloride (NED) to form a reddish purple azo dye.

Fig. 7. Transchem soil test kit and reagents

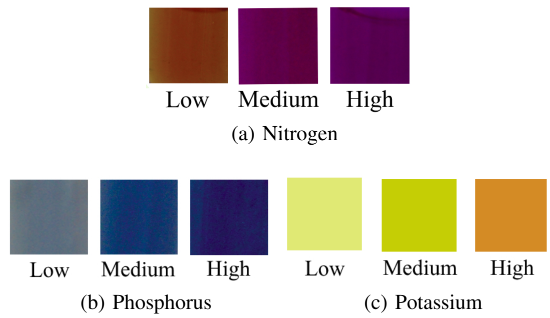


Fig. 8. Color variations from Low to High concentration for each nutrient

To determine phosphates, the extracted soil sample will be subjected to a color reaction with ammonium molybdate in the presence of suitable reducing agents to form a blue colored complex, the intensity of which is directly proportional to the concentration of phosphate in the solution.

To determine potassium, sample made from soil will be reacted with lithium dipicrylaminate resulting in a solid complex of potassium dipicrylaminate. Orange color complex will be dissolved using a suitable solvent. The intensity of the color is proportional to the amount of K^+ in the extract.

After preparation of the cassettes, the developed color is measured in the nutrient color sensor device.

C. Calibration and nutrient content estimation using Gaussian process regression

The Gaussian process model is a probability distribution over possible function space which fits the prior data points. It is particularly useful when the underlying function of the data is not known and only a few data points are available for training [8]. Gaussian process regression was chosen due to fact that the relationship of the color sensor readings and the nutrient concentration is not straightforward. Despite it can be determined using the Beer Lambert law and the sensitivity measurement for the RGB readings in the color sensor, it is quite complex and requires a lot of data. This relationship can be modeled more practically and accurately using the gaussian process regression model using a small dataset. The mean function $m(.)$ is used to estimate the unseen data, and the covariance $k(.,.)$ indicates the confidence level of the prediction made.

$$f(x) = GP(m(x), k(x, x')) \tag{1}$$

Kernel function is the main part of Gaussian process regression [9]. There are several kernels available for selection. Cross-validation method was used for the kernel model selection, where the Radial Basis function, rational quadratic, combination of several kernels with White kernel and constant kernel were evaluated with the training data. Rational quadratic kernel was chosen as the kernel function as it had the least mean squared error and more accurate model with the available data.

From the Beer-Lambert's law, it is known as that the relationship between the absorption and the concentration in the solution is non-linear and not periodic. Considering these, rational quadratic kernel is selected as the best kernel for the Gaussian process regression. The Rational Quadratic kernel is a scale mixture of RBF (Radial basis function) kernels with different characteristic length scales. It is parameterized by a length scale parameter l and a scale mixture parameter α. $d(.,.)$ is the Euclidean distance.

$$k(x, x') = \left(1 + \frac{d(x, x')^2}{2\alpha l^2}\right)^{-\alpha} \quad (2)$$

Rational quadratic kernel gave a more smooth and accurate behaviour with the appropriate length scale parameter which was carefully chosen using k-fold cross validation with different length scales. Since the relationship between the RGB colorspace and the nutrient concentrations needs to be estimated with a small dataset, estimating using a conventional regression method is challenging. However, Gaussian process regression provides a capability to interpret a function space suitable for the requirement by providing an appropriate covariance function with limited data. Ability to get the prediction confidence level is an added advantage using Gaussian Process regression.

D. Smartphone application

Smartphone application is the controller interface of the nutrient sensor device. The application is developed to obtain the color readings from the nutrient sensor device and predict the nutrient values using the pre-trained regression model. To establish a connection with the device, Bluetooth serial connection is used since it is easy to pair the device and connect to the existing Bluetooth radio of ESP32.

When connected to the device via Bluetooth, using the smartphone application as shown in Figure 9, reading of color can be triggered using the button, then a read signal is sent to the nutrient sensor device which starts reading the colors. The smartphone application listens to the Bluetooth serial port until data is received from the device. The color readings are sent to the smartphone application in JSON format, which contains 9 values.

Then a data model for the color readings, location data, timestamp and predicted results are stored in a geo-tagged database. This data is stored in a cloud database (Google Firebase and Firestore [10]) to retrieve to train and test the Gaussian process regression model and also to predict the nutrient concentrations. The trained models are stored and predictions are made using this pre-trained models with the nutrient sensor readings.

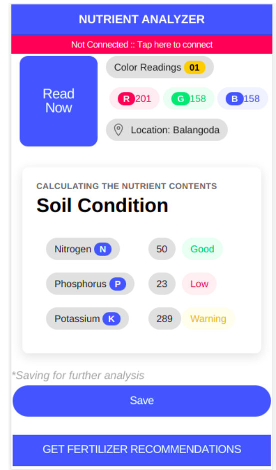


Fig. 9. Android application user interface: Read Now button is the trigger to start the reading upon Bluetooth connection. Output nutrient readings provided in the Soil Condition section

IV. Results

A. Phosphorous concentration measurement

To validate the estimation of soil nutrient levels using the proposed methodology, we used a chemical laboratory experiment with the required reagents and the prepared chemical samples containing fixed concentrations of soil.

Phosphorus was prepared in the initial testing and calibration. For this, 5 concentration samples were used as shown in Table I. The reagents were added as recommended and waited 10 minutes to develop the colors. The color is different in intensity as observed in Figure 10.

TABLE I
TRAINING AND TEST SAMPLE CONCENTRATIONS TAKEN FROM THE LAB EXPERIMENTS FOR PHOSPHOROUS

Concentration (ppm)	0	0.625	1.25	3.75	5
Trials	3	3	3	3	3

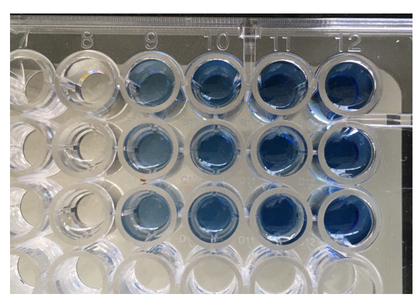


Fig. 10. Samples created from the different concentration levels, 0 ppm, 0.625 ppm, 1.25 ppm, 3.75 ppm, 5 ppm samples from left to right

Sample testing cassette shown in Figure 1 made of transparent acrylic was used to insert the colored solutions. The

15mm x 10mm x 2mm size was used to measure the color developed in the above process. 300 µL were inserted from each trial sample to cassettes using syringes.

K-Fold cross validation was used to train and validate the model with rational quadratic kernel [11] which gave a mean squared error of 0.32 when k=5. Table II shows the training iterations in the k-fold cross validation and the log marginal likelihood. Kernel length scale was chosen using the minimum negative log marginal likelihood value of the k-fold cross validation. Hence from the model 4, the length scale was 2564 which is used as the length scale of the GPR model.

TABLE II
K-FOLD CROSS VALIDATION RESULTS FOR K=5 SCENARIO

Model	Predicted Value	Actual Test Value	Standard deviation	Mean Squared Error	Log Marginal Likelihood
1	0.1868	0	0.722	0.039027	-29.038
	0.7431	0.625	0.7069		
	1.5113	1.25	0.54		
2	0.3119	0	0.9512	0.299823	-27.787
	4.353	3.75	0.2784		
	4.3374	5	0.4134		
3	0.9736	0.625	0.7325	0.27203	-28.644
	2.0339	1.25	0.8853		
	4.0331	3.75	0.3169		
4	-0.0058	0	0.4695	0.890985	-21.25
	4.8212	3.75	0.2216		
	3.7649	5	0.2141		
5	0.5568	0.625	0.7767	0.132632	-28.562
	1.2159	3.75	0.4887		
	4.3738	5	0.2628		

The trained Gaussian process regression models in k-fold cross validations were used to validate the test samples as shown in Figure 11. For each test value and predicted value in the scatter plot, it is evaluated whether the predictions are in the correct concentration region. From the test dataset, only one instance predicted an invalid region. In the agricultural usage of the nutrient concentration, it is mostly a regional reading which is required to determine the need for the fertilizer input. So low, sufficient and high regions can be correctly predicted 93% of the dataset.

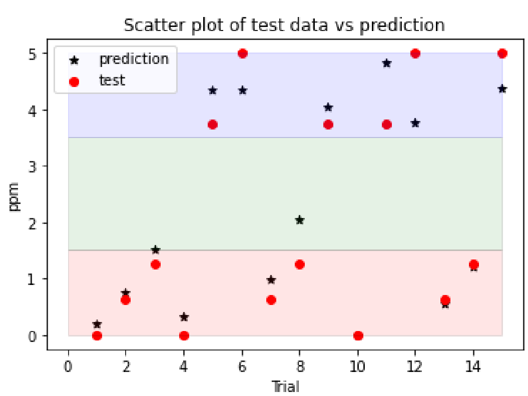


Fig. 11. Scatter plot of the test data and the model predictions. Red region represents the low level, green region shows the medium level and the blue region shows the high level

V. Conclusions

We have proposed an inexpensive and portable IoT device and a machine learning enabled mobile app that connects wirelessly with the device that can be used to easily estimate soil nutrient contents based on colorimetry. In this paper we have validated our approach using a set of solutions of known phosphorus concentrations but the setup generalizes to any nutrient that gives out a color change that depends on the concentration of the nutrient, after reacting with a given set of chemical reagents (soil test kit). As opposed to many methods that rely on comparing colors with the human eye, our approach provides a way to get consistent and accurate readings. Since we use Gaussian process regression to predict nutrient concentrations, it only needs a few data points for training and provides a principled way of producing the variance of estimates.

With the relevant data on the ideal nutrient concentrations for each crop and other related data for the successful crop yield, the soil nutrient reader device readings can be enhanced to a system which can provide standard fertilizer recommendations for each field.

The GPR model gives an average accuracy score of 0.88 for the training dataset. The nutrient concentration level can be predicted with a mean accuracy of 93%. This is sufficient to get a comprehensive understanding on the nutrient conditions in the soil which can help to take rapid actions towards fertilizer recommendations.

References

[1] R. G. Regalado and J. C. D. Cruz, "Soil ph and nutrient (nitrogen, phosphorus and potassium) analyzer using colorimetry," in *2016 IEEE Region 10 Conference (TENCON)*, pp. 2387–2391, IEEE, 2016.

[2] J. C. Puno, E. Sybingco, E. Dadios, I. Valenzuela, and J. Cuello, "Determination of soil nutrients and ph level using image processing and artificial neural network," in *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, pp. 1–6, IEEE, 2017.

[3] R. Madhumathi, T. Arumuganathan, R. Shruthi, and R. S. Iyer, "Soil nutrient analysis using colorimetry method," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pp. 252–256, IEEE, 2020.

[4] S. Agarwal, N. Bhangale, K. Dhanure, S. Gavhane, V. Chakkarwar, and M. Nagori, "Application of colorimetry to determine soil fertility through naive bayes classification algorithm," in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, 2018.

[5] A. Lazaro, M. Boada, R. Villarino, and D. Girbau, "Color measurement and analysis of fruit with a battery-less nfc sensor," *Sensors*, vol. 19, no. 7, p. 1741, 2019.

[6] Texas Advanced Optoelectronic Solutions Inc., "*TCS3472 COLOR LIGHT-TO-DIGITAL CONVERTER with IR FILTER*." "https://cdn-shop.adafruit.com/datasheets/TCS34725.pdf", TAOS135, 2012 [Online].

[7] Department of Agriculture, Ministry of Agriculture, Sri Lanka, *Site Specific Fertilizer Recommendation (SSFR) using soil test kit*, 2019.

[8] T. D. Ambegoda, J. N. Martel, J. Adamcik, M. Cook, and R. H. Hahnloser, "Estimation of z-thickness and xy-anisotropy of electron microscopy images using gaussian processes," *arXiv preprint arXiv:2002.00228*, 2020.

[9] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*, vol. 2. MIT press Cambridge, MA, 2006.

[10] Google Inc., "Cloud firestore — firebase documentation." https://firebase.google.com/docs/firestore, [Online].

[11] Scikit-learn developers, "1.7. gaussian processes." https://scikit-learn.org/stable/modules/gaussian_process.html, [Online].