Digital Soil Mapping using Machine Learning

Ashu Bansal, Riya Jain, Muskan Rastogi, Vimal Kumar

Department of Computer Science and Engineering, Meerut Institute of Engineering and Technology, Meerut,Uttar Pradesh, India

---

ABSTRACT

Agriculture is a non-technical sector where in technology can be incorporated for the betterment. Soil analysis is a method to analyse the available plant nutrients in the soil. Soil provides major nutrients to the plants. To create a prediction engine for most appropriate crop for a particular soil. It also determines the type of soil and its fertility. This work predicts the suitable crop and the fertility of a particular soil by analyzing the major and micro nutrients present in the soil. There are mainly three soil parameters that come into consideration when we have to predict the quality of the soil. This method suggests the soil fertility and suitable crop for a soil using Machine Learning Techniques.Our result gives the compatible crop for a particular soil sample by considering important soil parameters and by applying appropriate Machine Learning algorithm. Suitable crop for a particular sample is predicted on the basis of NPK factor, type of soil according to the pH level and soil fertility on the basis of major and micro nutrients with the maximum accuracy.

Keyword : Nutrients, parameters, Machine Learning, Accuracy.

1 INTRODUCTION

Agriculture is a non-technical sector where in technology can be used for the efficient management of soil. There is quick implementation and easy in adoption in the agricultural technology. Usually farmers used traditional method known as crop mutation after every subsequent crop yield [1]. In many countries, this traditional method is implemented where the change in crop is done after a loss in yield for cultivating the same crop continuously. This method helps the soil to regain the nutrients that were consumed by the crop previously and use the remaining nutrients for cultivating the new crop. Soil fertility is also maintained by this process. Farmer has to face a loss in yield when they come to know about condition of soil which is unfit to yield the particular crop [2]. About one financial year is important for a farmer to accept the loss in yield. Solution to the above stated problem is suggested using Machine Learning Techniques. There are mainly three soil parameters that becomes necessary in the prediction of soil fertility and suitable crop for a soil sample.

- Chemical Parameters
- Physical Parameters
- Biological Parameters Table 1 Soil Parameters

Parameters	Chemical	Description
Physical	Electrical Conductivity	Ions present in the soil sample are measured by the EC of soil. Conductivity of the soil increases when there is a movement of ions.
	Texture	Land consists of varieties of soil such as Clayey, Sandy, Layered, Semi-Layered etc.
	pH	It is a scale used to specify the acidic or basic character of an aqueous solution.
Chemical	Sulphur	It helps in the production of chlorophyll that is required for the photosynthesis.

	Phosphorous	It promotes the root growth of plants and make them to withstand low temperature.
	Potassium	It is a macronutrient present in the soil matter.
	Organic Carbon	It has an importantrole in the physical, chemicaland biological function of agricultural soils.
	Ferrous	Soil also consists of major iron content. Iron richsoil is acidic in nature.
	Zinc	It helps the plantsto produce chlorophyll.
	Boron	The reproductive and vegetative growth of plants isaffected by the deficiency of boron.
Biological	Micro- organism biomass of C and N	Microbes decompose soil organic matter releasing carbon dioxide and plant available nutrients.
	NaturalManure	Manure like vegetable waste and the animal excreta that saturates the organic matter insoil for plant growth.

A soil sample contains different types of elements that shows different behaviour with different crops. This table shows those components that are present in a sample.

There are the demerits of the abovetraditional method:

- Scope for redundancy
- Time lag
- Requires more human force
- Needs more labs

Soil Testing and analysis gives the accurate composition of the soil and the respective compatible crop. It directlybenefits a large number of the user (farmers or people associated with agriculture and farms). Machine Learning is always important for large dataset.Machine Learning Techniques make theprocessing flexible and automated usingalgorithms [3]. This project helpful to predict the fertility of soil by using some machine learning algorithms like Decision Tree Classifier, KNN Classifier and Random Forest Algorithm. Theclassification of soil according to thefertility also can be made easy by analysing the major and micro nutrients of the soil [4]. It takes less time to predict the fertility of soil than the traditional system as the machines worked faster and more efficient than the manual system. We can also note the fertility and the type of the soil very easily and efficiently and within less time. Our aim is to come up with an automatic soil testing system which not only will analyse the soil samples but also provide acceptable crop information at freeof cost and by consuming less time. This crop prediction is finished by not just considering the fertility of the soil but also by the type of a soil sample [5].

Using machine learning ideas, the handlingof multi-dimensional and heterogeneous information in dynamic settings can be performed.

- Easily identifies trend and patterns
- Fast processing and prediction inreal time
- Tasks are implementedautomatically easily
- Makes better decision

The contribution of this paper is constructed as follows:

Section 2 elaborates literature survey for research. Section 3 depicts themethodology and section 4 discusses the result of the proposed model and section 5 concludes the research paper by summarizing our work.

2 Related Work

This step involves breaking down the system into separate components to assess the situation, analysing the project's goals, breaking down what needs to be established, and attempting to engage clients to identify particular requirements [6].
Based on the nutrient present in the soil, the soil fertility would be predicted [7].
For crop analysis, we need to monitor various physical parameters such as Texture of soil, pH level or Electrical Conductivity etc.
Overall analysis of soil is being carried outin the project based on the respective parameters. Crop selection method has been developed for season wise cropprediction [8]. Therefore, based on Kharif (crops which are sown at the beginning of the rainy season, e.g., between April and May.), Rabi (crops that are sown at the end of monsoon or at the beginning of winter season, e.g., between September and October. These crops are known as monsoon crops.) and Zaid (short season between Kharif and Rabi season in the months of March to July) the seasonal crops will be predicted [9]. For prediction, they have compared and analysed differentalgorithms. One of the main factors that affect crop growth is texture of the soil.

Future vision & Scope

In the future we are expecting to work with the real time data/primary data. Currently we operate our research on secondary data due to some limitations and lack of resources. We are planning to deploy our algorithm to the cloud so that any device without prior training and testing of thedata can use it. We can introduce new machine learning algorithms and tactics to improve the accuracy of the testing Dataset As currently we have a stable maximum accuracy. We can wide spread our model suitable to various climatic conditions and zones so that it does not stick to a particular zone. We have planned to provide a hard copy of report or the hard copy of compatible crop to the farmers and to the users. This saves their time and money too.

3 Methodology

Sample of soil from SHC has been takenfrom secondary source to train the model. After, we removed the unwanted attributesthat don't contribute in the training model. Pre-process the data by adding newattributes to help in further processing."Soil type" is added on the basis of pHvalue of the soil. "Soil fertility" is addedon the basis of B, C ,N content. Afterward,we divided the data into 80-20 rule fortraining and testing of the data. In the nextstep we have used three ML models:Decision Tree Classifier, Random Forest,KNN to compare their respective accuracy. The model with better accuracy than othertwo's has been accepted for training.
Fig 3.1 shows the step by step flow chartof methodology accepted by our model.

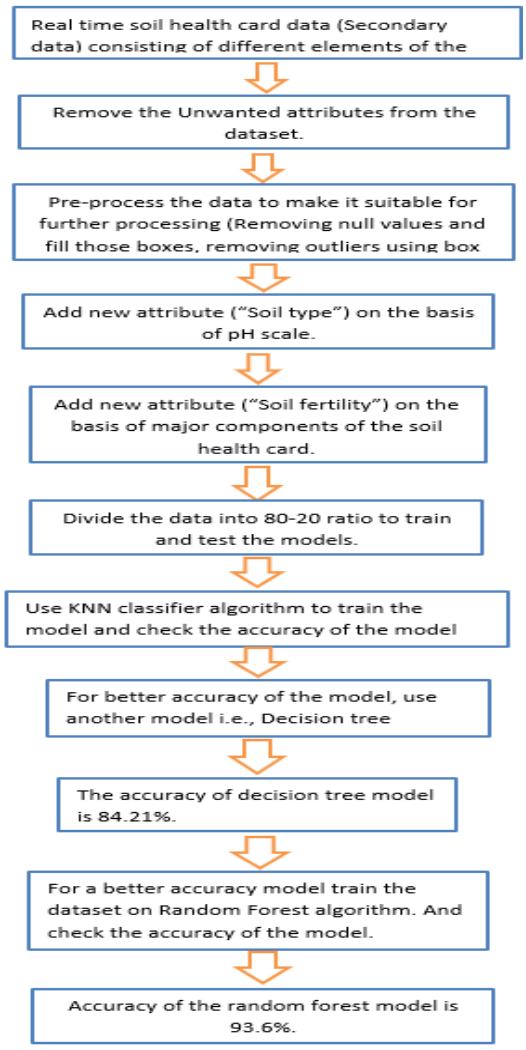

Figure 3.1: Flowchart of the Methodology

```
pH          float64
EC          float64
C           float64
P2O5        float64
K2O         float64
Sulphur     float64
Zinc        float64
Boron       float64
Ferrous     float64
dtype: object
```

Figure 3.2: Input Data Column

	pH	EC	C	P2O5	K2O	Sulphur	Zinc	Boron	Ferrous
count	753.000000	753.000000	753.000000	753.000000	752.000000	753.000000	753.000000	753.000000	753.000000
mean	7.635458	0.252337	0.406481	14.465206	173.187500	7.237009	5.261116	1.866939	2.695631
std	0.273675	0.518750	0.156509	1.653473	73.670575	7.705094	6.550918	3.961555	2.000558
min	6.000000	0.080000	0.000000	3.000000	49.000000	0.000000	0.000000	0.001000	0.000000
25%	7.500000	0.150000	0.300000	13.500000	120.000000	3.100000	1.400000	0.001000	2.000000
50%	7.600000	0.190000	0.400000	14.500000	160.000000	5.800000	2.600000	0.904000	2.500000
75%	7.800000	0.260000	0.500000	15.500000	210.000000	8.600000	7.000000	2.131000	3.000000
max	9.300000	13.900000	1.050000	22.000000	570.000000	98.130000	67.700000	65.000000	30.910000

Figure 3.3: Data after Processing

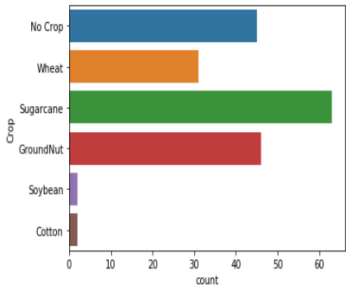

Figure 3.4: Data ML Model

4 Results:

Gather the data recorded in Krishi Vigyan Kendra for the soil health report of thearea[10]. Check the attributes suitable for the processing of the soil testing. On the basis of selected attributes removesunwanted attributes and add some attributes like (Soil type, soil fertility) to the existing data. Remove the outliers using box plot method. Divide the data into 80-20 to train and test the ML model respectively.

Starting with the KNN model, train ourdataset on this model and we get the accuracy of the model. We have accuracy of 64.08% (where n=6). For a better accuracy we train our model under Decision tree classifier model, and get the accuracy of 84.21%. To get a betteraccuracy of the model we train the model with Random Forest algorithm and get the accuracy of 93.6%.

The accuracy of the model signifies that for a set of well define attributes our modelcan predict a crop that fits 93.6% to the fertility of the soil.

4.1 Decision Tree Classifier:

Decision Tree is a type of supervised learning used for classification and regression problems. It aims to build a model that predicts the value of a target variable by learning simple decision rulesinferred from data nodes.

	precision	recall	f1-score	support
Cotton	0.00	0.00	0.00	0
GroundNut	1.00	0.78	0.88	9
No Crop	0.80	0.89	0.84	9
Soybean	0.50	1.00	0.67	1
Sugarcane	0.92	0.92	0.92	13
Wheat	1.00	0.83	0.91	6
accuracy			0.87	38
macro avg	0.70	0.74	0.70	38
weighted avg	0.91	0.87	0.88	38

Accuracy_score
0.868421052631579

Figure 1: Working of Decision TreeClassifier

4.2 KNN CLASSIFIER

KNN stands for K-Nearest Neighbours andit is based on Supervised Learning technique. It is one of the simplest algorithms of Machine Learning. It works on the similarity measure between the input data and the available data and put the new data into the category which is similar to the available category. This algorithm is mostly used for the Classification problems but either be used for Regression or for Classification. considered were KNN Classifier, Decision Tree, K nearest neighbour and Random Forest and among all four algorithms, the accuracy rate for Random Forest was high.Accuracy for each algorithm is shown in Table 2 below.

Table 2 Accuracy of different algorithms

Algorithms	Accuracy
K-Nearest Neighbour	62.46%
KNN Classifier	64.08%
Decision Tree	84.21%
Random Forest	93.6%

	precision	recall	f1-score	support
Cotton	0.00	0.00	0.00	0
GroundNut	0.67	0.89	0.76	9
No Crop	0.75	0.67	0.71	9
Soybean	0.00	0.00	0.00	1
Sugarcane	0.69	0.69	0.69	13
Wheat	0.67	0.33	0.44	6
accuracy			0.66	38
macro avg	0.46	0.43	0.43	38
weighted avg	0.68	0.66	0.65	38

Accuracy_score
0.6578947368421053

Figure 2: Working of KNN Classifier

4.3 RANDOM FOREST

We compared our data model accuracy Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. Both Classification and Regression problems can be solved by Random Forest. It is a process of combining multiple classifiers to solve a common problem and to improve the performance of the model which is the concept of ensemble learning.

Classification report

	precision	recall	f1-score	support
GroundNut	1.00	0.89	0.94	9
No Crop	0.80	0.89	0.84	9
Soybean	0.00	0.00	0.00	1
Sugarcane	0.93	1.00	0.96	13
Wheat	0.83	0.83	0.83	6
accuracy			0.89	38
macro avg	0.71	0.72	0.72	38
weighted avg	0.88	0.89	0.88	38

Accuracy_score
0.8947368421052632

Figure 3: Working of Random Forest

Different algorithms are to be compared. Different algorithms gave distinct results on the same dataset. The algorithm with the pre-published Research paper onsoil testing and prediction [11].

Table 3 Comparison between Methods

Algorithms	Published Method	Our model Method
Decision Tree Classifier	61.5%	84.21%
Random Forest	72.74%	93.6%

Graph is plotted by using the above result.

Graph Comparison

```
In [52]: import matplotlib.pyplot as plt
         import numpy as np

In [51]: old_readings=[61.5,71.24]
         new_readings=[84.21,93.6]

In [59]: barWidth = 0.25
         fig = plt.subplots(figsize =(6, 6))
         br1 = np.arange(len(old_readings))
         br2 = [x + barWidth for x in br1]

         plt.bar(br1, old_readings, color ='r', width = barWidth,
                 edgecolor ='grey', label ='old_readings')
         plt.bar(br2, new_readings, color ='g', width = barWidth,
                 edgecolor ='grey', label ='new_readings')

         plt.xlabel('Algorithms', fontweight ='bold', fontsize = 15)
         plt.ylabel('Accuracy', fontweight ='bold', fontsize = 15)

         plt.xticks([r + barWidth for r in range(len(old_readings))],
                 ['Decision Tree Classifier', 'Random Forest' ])
         plt.legend()
         plt.show()
```

Figure 4: Working of Graph Comparison

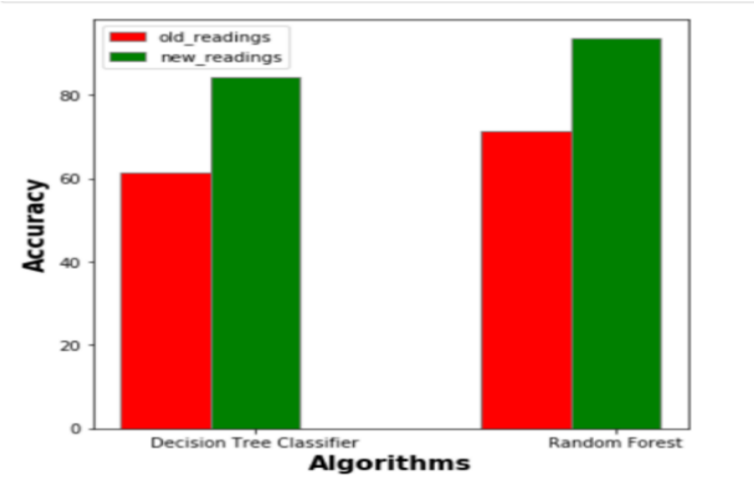

Figure 5: Graph Comparison

This figure shows the accuracy comparison of our model that is named as new_readings and we compared thosevalues by another research paper named asold_readings in our paper that is already published.

5 Conclusion

In this project analysis of soil based on major and minor nutrients present in the soil has been proposed using Machine Learning Techniques. The project has high efficiency and accuracy in fetching the real time dataset of soil components. The project will assist the farmers in increasing the agriculture yield and take efficient care of crop production as the stick will always provide helping hand to farmers for getting accurate live feed of soil fertility and the type of the soil upto93.6% accurate results. The project proposes a wise agricultural model in integration with Machine Learning. Machine Learning have always mattered in agriculture domain. It is really challenging task because of highly localized nature of agriculture information specifically distinct conditions. We have used number of algorithms such as Decision Tree Classifier, KNN Classifier and Random Forest. After testing each algorithms we get accurate result with 93.6% by Random Forest. The complete real-time and historical environment information is expected to help for betterment.

References

1. T. Venkat Narayana Rao, "Prediction of Soil Quality using Machine Learning Techniques", International Journal of Scientific& Technology Research Volume 8 (2019), ISSN 2277-8616

2. P. Shubham, R. Prem, S. Swami, P. Sandip, "Soil Analysis and Crop Prediction", International Journal of Scientific Research in Scienceand Technology (2020), ISSN: 2396-6011, Vol. 7, Issue 4, Pagenumber: 117-123

3. K. Vishal, K. Raushan, K. Shubham, Ajinkya, Prof. P. P. Jorvekar, "Agriculture soil Analysis for Suitable Crop prediction".

4. Kazheen I T, Adnan M A, Dilovan A Z (2021), "Data Mining Classification Algorithms for Analyzing Soil Data", Asian Journal of Research in ComputerScience, 8(2): 17-28

5. R. Jayalakshmi, M. Savitha Devi, "Predictive Model Construction for Prediction of Soil Fertility using Decision Tree Machine Learning Algorithm".

6. P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining", IJARCSSE, vol. 6, Issue 5, 2016.

7. Sabri Arik, Tingwen Huang, Weng Kin Lai, Qingshan Liu, "Soil Property Prediction", An Extreme Learning Machine Approach|| Springer, vol. 3, Issue 4,666-680, 2015.

8. Sadia Afrin, Abu TalhaKhan, "Analysis of Soil Properties and Climatic Data to Predict Crop Yields " (2018 IEEE ICIS 2018).

9. Singh "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique"2015 IEEE, International Conference on Smart Technologies and Management for Computing,Communication, Controls, Energy and Materials (ICSTM).

10. S.K.D Krishi Vigyan Kendra, Baghra, Muzaffarnagar, established by Shri Shri 1008 Veetrag swami Kalyan Dev ji Maharaj, December (1995)

11. G. Jay, I. Anurag, G. Shailesh, A. Vahida, "Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction","InternationalJournal of Computer Science", Issue 9, June (2012)

12. Abishek.B, AkashEswar"Prediction of Effective Rainfall and Crop Water Needs using Data Mining Techniques" (2017 IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development (TIAR 2017)).

13. Rajkomar, A., Lingam, S., Taylor, A. G., Blum, M., & Mongan, J. (2017). Highthroughput classification of radiographs using deepconvolutional neural networks. Journal of digital imaging, 30(1), 95-101.

14. Tensorflow. (2019, July 10). Tensorflow/tensorflow. Retrieved August 04, 2019, from https://github.com/tensorflow/tensorflow/tree/master/tensorflow/exam ples/label _image

15. Dyrmann, M., Karstoft, H., & Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. Biosystems Engineering, 151, 72- 80.