International Journal of Advanced Engineering and Technology
ISSN: 2456-7655; Impact Factor: RJIF 5.54
Received: 16-04-2019; Accepted: 19-05-2019
www.newengineeringjournal.com
Volume 3; Issue 3; July 2019; Page No. 42-46

A study on prediction of soil properties content based on machine learning models

Charu Gitey[1], Dr. Kamlesh Namdev[2]
[1] M. Tech Scholar, Department of Computer Science & Engineering, SIRT- Excellence, Bhopal, Madhya Pradesh, India
[2] Associate Professor, Department of Computer Science & Engineering, SIRT- Excellence, Bhopal, Madhya Pradesh, India

Abstract
The Indian economy is dependent on the agricultural sector. Although focused on industrialization, agriculture remains an important sector of the Indian economy, both in terms of contribution to gross domestic product (GDP) and jobs for millions of people across the country. One of the key factors for productive agriculture is soil. The purpose of the work is to predict the type of terrain using data mining classification methods. Agricultural properties and soil ownership play a crucial role in agricultural decision-making. Now-a-days there is a need to study the nutrient status in lower horizons of the soil. Soil testing has played historical role in evaluating soil fertility maintenance and in sustainable agriculture. Soil testing shall also play its crucial role in precision agriculture. At present there is a need to develop basic inventory as per soil test basis and necessary information has to be built into the system for translating the results of soil test to achieve the crop production goal in new era. To achieve this goal Machine Learning approach is used for predicting the soil properties.

Keywords: yield prediction, data mining, soil analysis, machine learning

1. Introduction
Soil is very important for plant life. It consists of solids (minerals and organic matter), liquids (water and solutes) and gases (mainly oxygen and carbon dioxide) and contains living organisms. All these elements provide their physical and chemical properties. To maintain fertility, achieve better yield and protect the environment, it is necessary to nurture the soil properly. On the other hand, soil tests are essential to manage it properly. A soil test is the study of a soil sample to discover an additional substance, its composition and various attributes. As a general rule, soil tests are performed to determine the wealth and indicate the gaps to be corrected [3]. The analysis of soil nutrients is very useful for the farmer in determining the type of yield to be grown in a particular soil condition.
Soil fertility, which refers to the intrinsic capacity of the soil to provide essential nutrients to plants in sufficient and adequate proportions for optimal growth, is one of the key elements for determining soil productivity. The management of Indian soil fertility requires sustainable high-level production to produce adequate food for the growing population. Good soil fertility management requires careful identification of the limits of current nutritional deficiencies and monitoring of changes in soil fertility to predict their shortage. These gaps must be mitigated by sound and best practices in terms of nutrients, water, plants and energy for soil management, in order to maintain food production at a reasonable level to ensure high productivity at the same time. Future [4-6]. Therefore, managing soil fertility at optimal levels is one of the key factors for achieving high and sustainable productivity.
Soil fertility is one of the most important factors in crop control. The macronutrients (N, P, K) and the micronutrients (Zn, Cu, Fe and Mn) are important soil components that control their fertility. The characterization of soil in relation to the assessment of the fertility status of an area or region is important in the context of sustainable agricultural production. Due to the unbalanced and inadequate use of fertilizers and the low efficiency of other inputs, the efficiency of the reaction (production) of chemical fertilizers has significantly decreased in recent years with intensive cultivation. Fluctuations in nutrient intake are a natural phenomenon and some of them may be sufficient while others are inadequate [7].
In this work, data mining classification methods are used to study soil nutrients. Data mining involves extracting information from a data set and transforming it into a structure that is understandable for future use. Various data collection methods are available for the field research sector. Classification is one of the data mining techniques that automatically create a model of classes from a set of records that contains class labels. Popular classification techniques include decision trees, neural networks, k-nearest neighbour, SVM, and Naïve Bayesian classifier etc [8].

2. Analysis of Soil
Soils are formed by the combination of weathered rock materials with humus. We also know that soils supplies water and nutrients to flora. Soils also protect and purify rain water, pests and wildlife habitation. A little bit of hydroponics (soilless agriculture) are developed throughout the globe, but still their percentage is only 10.6 of total soils under conservation. The country prosperity is depending upon soils of that country [9].
Soil analysis and their classification is very critical because in the entire globe the types of the soils are same, but their analysis results may vary from location to location depending on various characteristics of soils. When analyzing the soils, we should consider the fundamental substantial, organic and compound properties of soils. The classification of soils deals with the methodical cataloging of soils depends on their individual characteristics, as well as decisive factor that dictate choice in use. Classification of Soils is one of the challenging areas in data mining and

machine learning, In classification of soils can be started from the viewpoint of soils as a matter and soil as a resource [10-12].

A. Types of Soil

The following are various types of soils in India.

Alluvial Soils: These soils are produced by the deposition of sediments carried by rivers. These soils are highly loaded with humus and very much productive. They are found in Great Northern plain, lower valleys of Narmada and Tapti and Northern Gujarat. These soils are transformed year by year.

Black Soils: The basic material of black soils volcanic rocks and lava-flow. These soils are spread over Deccan Lava Tract which includes parts of Maharashtra, Chhattisgarh, Madhya Pradesh, Gujarat, Andhra Pradesh and Tamil Nadu. These soils have huge amount of clay, it may be more than 62%. These soils have chemical composition of alumina, iron oxide, magnesium carbonates and lime and also Potash but are short of in Phosphorus, Nitrogen and Organic matter.

Red Soils: The color of these soils is due to heavy presentation of iron oxide. Acidic content in these soils are high, and phosphates and nitrogen is less. These soils are unable to retain moisture and water. These soils are formed due to slow breaking of metamorphic rocks and crystalline. They spread over the whole of Tamil Nadu, Andhra Pradesh, Chhattisgarh, Karnataka, Maharashtra and parts of Orissa.

Literate Soils: These soils are formed where heavy temperatures and heavy rainfalls occurred alternatively, normally these soils are final creation of weather. These soils contain good composition of bauxite or ferric oxides and less composition of potash and nitrogen. These are spread around in Kerala, Tamil Nadu, Maharashtra, Chhattisgarh and hilly areas of Orissa and Assam.

Mountain Soils: These soils are formed at hill slopes by gathering of organic matter derived from forest and woodlands. They are found in Himalayan region and also appeared in other regions according to altitude. These soils characteristics are based on climate, ground configuration and on parent rocks.

Desert Soils: In the desert regions of Rajasthan, soils don't seem to be well developed. As evaporation is in way over rain, the soil includes a high salt, alkaline content. The nature of the soil present in these areas are sandy and contains less organic components.

3. Properties of Soil

Soil may be an important element of the earth system, functioning not just for the assembly of food, fodder and fiber however additionally within the maintenance of native, regional and global environmental quality.

Soil fertility that refers to the inherent capability of a soil to produce essential nutrients to plants in adequate and right proportion for optimum growth, is one in all the key parts to see soil productivity. Management of the fertility of Indian soils demands for sustainable production at a high level to produce adequate food to feed its burgeoning population. Proper management of soil fertility requires careful determination of the limits of current nutritional deficiencies and monitoring of changes in soil fertility to predict soil nutrient deficiency. Solid, proven nutrient, water, plant and soil management solutions need to be addressed to maintain food production at a reasonable level to ensure high productivity in the future [13-15]. Therefore, managing soil fertility for optimal nutrient management is one of the key factors for high and sustainable productivity.

Soil fertility is one of the important factors influencing crops. Macronutrients (N, P, K) and micronutrients (Zn, Cu, Fe and Mn) are important soil elements that affect fertility. The characterization of soil in relation to the assessment of the fertility status of an area or region is important in the context of sustainable agricultural production. Due to the unbalanced and inadequate use of fertilizers and the low efficiency of other raw materials, the efficiency of the reaction (nutritional value) of chemical fertilizers has significantly decreased in recent years with intensive cultivation. Variation in nutrient intake is a natural phenomenon and some may be sufficient if there are other deficiencies [16].

Nitrogen being an essential constituent of protein is a vitally important plant nutrient. Nitrogen is one of the 17 essential plant nutrients indispensable for growth and development of crop plants. According to Katyal (2016), currently N deficiency affected 99% of the soils and crops grown on them invariably show deficiency symptoms in almost all the fields where it is not applied. An adequate supply of nitrogen is generally associated with vigrous vegetative growth of plants and deep green colour of leaves, as it is a constituent of chlorophyll as well as DNA and RNA.

Phosphorus stands next in importance among the essential nutrients and the soils of the region are low in P content. Phosphorus, amongst NPK can be called “The master key” with respect to the yield and quality of crops. It is useful for the absorption of different nutrients as the most prominent effect of P has been observed on the root system of plants as it promotes the formation of lateral and fibrous roots and a constitute of ATP and ADP, the energy supplying sources of plant.

Potassium is essential for various metabolic activities of living cell, transformation of carbohydrates, reduction of nitrates, synthesis of protein and normal meristematic activities where it acts as a catalyst or as a co-factor in enzymatic reaction of living cells. It has been suggested that potassium may also affect photosynthesis maintenance of turgor in plant cells as well as formation of oil and imparting disease and insect resistance. Its negligence in use has provoked widening of its deficiency in soils across length and breadth of the country. Laha *et al*. (2003) have also confirmed depletion of soil reserves [20-25].

Sulphur (S) has been recognized as the fourth important plant nutrient after nitrogen (N), phosphorus (P) and potassium (K). It is one of the essential elements for plant growth and activated many enzymes which are involved in photosynthesis and nitrogen fixation are attributed to the type of sulphur linkage present. Sulphur has been found to help the synthesis of amino acid and hence increase protein content of plants, boosts the oil content. Heavy removal of sulphur by high yielding varieties particularly oil seed crops, intensive cropping with high S requiring crops and losses of sulphur from soil through leaching and erosion causes sulphur deficiency in soil. It is also becoming deficient in soil due to continuous use of sulphur free fertilizers. The deficiency of S in general causes accumulation of nitrates and amides in plants failed to utilize N, entire plant becomes yellow, growth is stunted and spindly [26].

Indian soils are generally not very fertile because they are continuously impoverished in finite nutrients due to their

endless cultivation for centuries without a continuous substitution of nutrients. The soil analysis summaries indicate that 98% of the Indian land in the P available has a medium-low value and 60% of the land in K is from low to medium, while N is still generally insufficient. In the last fifteen years, in an intensive production system, there has been a phenomenal increase in sulfur deficiency using sulfur-free sulfur-free fertilizers.

Performance losses have been reported, even if the NPK fertilizer is used consistently. The decrease in yield is usually due to the lack of secondary nutrients and micronutrients. According to Rattan *et al*. (2009), more than 2.5 lakh soil samples were analyzed as part of a global micronutrient research project in India in 20 states in the country. It was found that samples 49, 33, 13 and 7% of the zinc, boron - samples had an iron or molybdenum deficiency. In Madhya Pradesh, zinc deficiency was observed in about 58% of soil samples, and in some states and crops, deficiencies in B and Mo also limit plant production [27].

Intensive cultivated land is depleted of available nutrients, in particular secondary nutrients and micronutrients. Therefore, an assessment must be made of the fertility status of intensively cultivated land with high yield crops. After the soil is detected, samples of composite soil are usually collected in fields without geographical references. The results of these soil tests are not useful for site specific recommendations and subsequent monitoring. The available nutritional status of a soil using a global positioning system (GPS) helps to formulate site-specific recommendations for balanced fertilizers and to temporarily and temporarily understand soil fertility status [28].

4. Related Work

India is an emerging economic power with vast human and natural resources, and a huge knowledge base. Ensuring food security is a key challenge facing our society. So to improve agricultural practices and the core of precision agriculture is to manage agricultural practices such as to accurately predicting responses of crop growth. Agriculture is the predominant occupation in India, accounting for about 52% of employment [1]. Agricultural soil quality includes the soil properties, particularly those related to organic matter such as N, C, pH, Mg Ca and K [2] and essential plant nutrients, K+, Ca2+, Mg2+ and NH4+ and detrimental elements, Na+, H+ and Al+3 [3]. Crop response to fertilization recommended micronutrient for plant growth are P, k, Mn, Zn and Cu [4]. The native productive capacity of soil is dependent on basic chemical and physical properties and predictions are possible [5, 6]. So the main aim is to analysis, adjustment, and establishment of crops growth factors and soil properties' parameters.

Bhuyar *et al*. [1] analyzed J48 classification algorithm in high accuracy for predict the soil fertility rate. In4 investigated the uses of various DM techniques for knowledge discovery in agriculture sector and introduced different exhibits for knowledge discovery in the form of Association Rules, Clustering, Classification and Correlation.

Suman *et al*. [4] proposed data mining techniques such as naïve bayes, k-nn and J48 for predicting soil fertility.

Ramesh *et al*. [5] predicted crop yield by using multiple linear regression technique. For this classification or prediction relationship among variables are analyzed. And further k-mean clustering is used to cluster and classify data on the basis of rainfall variables.

Tsai *et al*. [6] analyzed the vegetative factors of landslides in the Shimen reservoir watershed in northern Taiwan. Decision tree, Bayesian Network data mining techniques and the non-linear approaches were implemented. Optimization based Bayesian Network approach was considered as better than non-linear.

Tittonell *et al*. [7] analyzed the virtual significance of soil fertility and the crop management factors to predict the maize yields and in determining the yield variability and the gap between farmers. Classification and regression tree analysis was used to predict the result.

Bindraban *et al*. [8] investigated two comprehensive methods to calculate the production related yield gap and a soil fertility related nutrient balance. The methodology allows knowledge from micro-scale to higher-scale levels and determines land quality.

Gholap *et al*. [9] predicted soil attributes and analyzed soil data using classification techniques. Soil properties such as pH value, Electrical Conductivity (EC), Potassium, Iron, Copper, etc. were classified using classification algorithms like Naïve Bayes, J48 and JRip. Among the algorithms, J48 was considered as simple classifier and produced better result.

Majumdar *et al*. [10] prosed a crop production technique that find the optimal parameters that gives minimum error with maximum crop production prediction. For this PAM, CLARA, DBSCAN and Multiple Linear Regression are used for classification. For that climatic condition, crop yield are analysed.

Shruti Kulkarni *et al*. [11] designed a data-driven model that learns by historic soil as well as rainfall data to analyze and predict crop yield over seasons in several districts, has been developed. For this study, a particular crop, Rice is considered. The designed hybrid neural network model identifies optimal combinations of soil parameters and blends it with the rainfall pattern in a selected region to evolve the expectable crop yield.

B. P. Ganasri *et al*. [12] proposed a data mining techniques that uses physical and water properties of areas present in Netravati river basin in Karnataka State, India. For prediction of soil properties neural network based on Radial Basis Function (RBF) are used and error estimated was 1.084%, 1.09%.

Sulaxana *et al*. [13] designed a soil sensor and analyze the errors in the prediction of a soil nutrient. The technique can predict soil urea content and is based on multivariate analysis using the PLSR (Partial Least Square Regression) mathematical tool. The results obtained show that the percentage error in prediction of urea is within the tolerable limits of +/-5% of the actual value, when other soil nutrient concentrations are varied below and above their normal values.

Hongmei *et al*. [14] studied topography and soil properties in Baoji which were selected to establish the machine learning model to predict the soil moisture. The prediction was carried out by three models, support vector machine (SVM), random forest (RF) and back-propagation neural network (BPNN). The results showed that the RMSE of SVM were 7.521 and 8.011 respectively, while the RF were 10.759 and 11.042, and of BPNN were 12.147 and 11.165.

Singhatiya [15] discussed the need to study the nutrient status in lower horizons of the soil. Soil testing has played

historical role in evaluating soil fertility maintenance and in sustainable agriculture. Soil testing shall also play its crucial role in precision agriculture. At present there is a need to develop basic inventory as per soil test basis and necessary information has to be built into the system for translating the results of soil test to achieve the crop production goal in new era. To achieve this goal artificial intelligence approach is used for predicting the soil properties. In this paper for analysing these properties support vector regression (SVR), ensembled regression (ER) and neural network (NN) are used. The performance is evaluated with respect to MSE and RMSE and it is observed that ER outperforms better with respect to SVR and NN.

5. Machine learning models

Machine learning is to learn the inherent regularity information in the data through the computer, gain new experience and knowledge, in order to improve the intelligence of the computer, so that the computer can make decisions like humans, the commonly used machine learning algorithms are decision tree, random forest, neural network, support vector machine, etc.

A. BP Neural Network

Artificial neural network technology is a mathematical model of distributed parallel information processing algorithms established by simulating the neural network behavior characteristics of the human brain. The BP neural network was first proposed by a scientist led by Rumelhart and McClelland to train according to the error reverse propagation algorithm. The multi-layer feed-forward neural network doesn't need to determine the mathematical equation of the mapping relationship between input and output in advance, but only learns certain rules through its own training, and the result that is closest to the expected output value is given when the input value is given [2]. The basic idea of the BP neural network is: if the desired output is not propagated forward using the weighted threshold, the back propagation is repeated, the weights and thresholds of the nodes are iteratively modified, and the cost function is gradually reduced until the preset value is reached. It is generally required that the cost function be smaller than a relatively small positive number or iteration and no longer be reduced, but rather iteratively oscillated until the training of the BP network and the determination of the mapping relationship between the input and output are completed [4]. In short, adjusting the weights minimizes the total network error.

B. RF

The random forest algorithm is a method proposed by Leo Breiman that uses multiple tree classifiers for classification and prediction. The random forest algorithm can be used to deal with problems such as regression, classification, clustering, and survival analysis; when used for classification or regression problems, its main idea was to resample by self-help method to generate a large number of tree regressors or classifiers. The basic principles of the random forest regression algorithm are: (1) Using the bootstrap method to select "n" sample training sets randomly from the original sample to form "n" decision trees (ntrees), each sample not being drawn out-of-bag (OOB) as a random forest test sample; (2) From the explanatory variables, extract the "m" explanatory variables that are most effective for partitioning the data, and determine the value according to the minimum prediction error principle of the bag. (3) Integrate the generated multiple regression trees into a random forest, and select the average of all the prediction values of the decision tree as the final result.

C. SVM

Support Vector Machine (SVM) is a kind of machine learning method that Vapnik *et al*. proposed in the 1990s and has a strict mathematical derivation and a solid theoretical foundation. Because of its unique advantages in computer image processing [5], nonlinear modeling and prediction, and optimization control, is considered as a new research attention following neural network learning [11]. Support Vector Regression (SVR) is a statistical learning-based VC dimension theory and structural risk minimization principle, it acquires the best learning model from the existing sample information. The core idea is to obtain the complicated nonlinear mapping relationship between the dependent variable and the independent variable learning through the sample information, is based on Mercer's kernel expansion theorem, the low-dimensional sample space is mapped to a high-dimensional and even infinite-dimensional feature space through the nonlinear mapping to make it in the feature space, which can be used to solve the problems of high regression in low-dimensional sample spaces [15].

6. Conclusion

This paper focuses on analysis of the key properties of soil, such as organic matter, plant nutrients, micronutrients that influence crop growth and identify the appropriate percentage ratio between these crops. properties that use supervised. Although these parameters can be measured directly, their measurement is difficult and expensive. The machine learning algorithms are trained on the basis of nutritional scale of different soil samples and predict the future crop production for future.

In conclusion, the result of this study showed that training is very important in increasing the model accuracy of one region and result in the form of a guide to recognizing soil properties relevant to plant growth and protection.

The present dataset gives the fertility status only in surface horizon (0-15 cm). Further there is a need to study the nutrient status in lower horizons of the soil. Soil testing has played historical role in evaluating soil fertility maintenance and in sustainable agriculture. Soil testing shall also play its crucial role in precision agriculture. At present there is a need to develop basic inventory as per soil test basis and necessary information has to be built into the system for translating the results of soil test to achieve the crop production goal in new era.

7. References

1. Bhuyar V. Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad District. International Journal of Emerging Trends and Technology in Computer Science. 2014; 3(2):200-3.
2. Shivnath Ghosh SS. A Review on Soil Property Detection using Machine Learning Approach", International Journal Online of Science. 2018; 4:8. Retrieved from http://ijoscience.com/ojsscience/index.

php/ojsscience/article/view/152. Doi : https://doi.org/ 10.24113/ijoscience.v4i8.152
3. Fathima NG, Geetha R. Agriculture crop pattern using data mining techniques. International Journal of Advanced Research in Computer Science and Software Engineering. 2014; 4(5):781-6.
4. Suman Naib BB. Soil classification and fertilizer recom¬mendation using WEKA. International Journal of Comput¬er Science and Management Studies. 2013; 13(5):142-6.
5. Ramesh D, Vardhan VB. Data mining techniques and ap¬plications to agricultural yield data. International Journal of Advanced Research in Computer and Communication Engineering. 2013; 2(9):3477-80.
6. Tsai F, Lai JS, Chen WW, Lin TH. Analysis of topographic and vegetative factors with data mining for landslide verification. Ecological Engineering. 2013; 61:669-77.
7. Tittonell P, Shephered KD, Vanlauwe B, Giller KE. Unravel¬ling the effects of soil and crop management on maize pro¬ductivity in small holder agricultural systems of Western Kenya - An application of classification and regression tree analysis. Agriculture, Ecosystems and Environment. 2008; 123(1-3):137-50.
8. Bindraban PS, Stroorvofel JJ, Jansen DM, Vlaming J, Groot JJR. Land quality indicators for suitable land management: Proposed methods for yield gap and soil nutrient balance. Agriculture, Ecosystems and Environment. 2000; 81:103-12.
9. Gholap J, Lngole A, Gohil J, Shailesh, Attar V. Soil data analysis using classification techniques and soil attribute prediction. 2012; 9(3):1-4.
10. Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki. Analysis of agriculture data using data mining techniques: application of big data", Journal of Big Data, 2017.
11. Shruti Kulkarni, Shah Nawaz Mandal, Srivatsa Sharma G. Predictive Analysis to Improve Crop Yield using a Neural Network Model, IEEE, 2018.
12. Ganasri BP, Dwarakish GS, Assessment of spatial variability of soil physical and hydraulic properties in Netravati catchment, Karnataka State, India, International Conference on Technologies for Sustainable Development, IEEE, 2015.
13. Sulaxana R. Vernekar, Ingrid Anne P. Nazareth, Jivan Parab S, Gourish Naik M. Error analysis in soil urea prediction based on RF spectroscopy, IEEE International Conference on Advances in Electronics, Communication and Computer Technology (ICAECCT), 2016.
14. Hongmei Nie, Lianan Yang, Xinyao Li, Li Ren; Jinhao Xu, Yongtao Feng. Spatial Prediction of Soil Moisture Content in Winter Wheat Based on Machine Learning Model, International Conference on Geoinformatics, IEEE, 2018.
15. Smriti Singhatiya, Dr. Shivnath Ghosh. Performance Evaluation of Artificial Intelligence on Soil Property Detection, International Journal Online of Science. 2018; 4:10. Retrieved from http://ijoscience.com/ ojsscience/index.php/ojsscience/article/view/166. DOI: https://doi.org/10.24113/ijoscience.v4i10.166