ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

Prediction of Suitable Crops for Better Production Using Machine Learning

Nithya P[1], Dr.A.M.Kalpana[2]

Assistant Professor, Department of Computer Engineering, Government College of Engineering, Salem, India[1]

Professor and Head, Department of Computer Engineering, Government College of Engineering, Salem, India [2]

ABSTRACT: In the field of agriculture, there is a problem in evaluating the production accuracy due to involvement of enormous crops. This problem can be overcome by effective machine learning(ML) method. Previously, some methods were utilized to evaluate the accuracy in better way but still, they are having some limitations in it. As a result, there are scarcity in efficient techniques and it will not yield required information to the users. This makes a motivation and tends to propose an effective and vigorous architecture named as enhanced back propagation model in neural network to predict the best suitable crops for agriculture. For evaluation propose, the parameters such as: the root mean square error (RMS), root relative square error (RRSE), normalized mean absolute error (MAE), and statistical measures are noticed.

KEYWORDS: machine learning, prediction, neural network, crops.

I. INTRODUCTION

Among many cultures in India, agriculture is an important one and it is followed from ancient times. In the stone age, the peoples used their land to produce the crops and hence the products are very organic. Apart from humans, animals and birds also had been a part of agriculture [1]. Prediction of crops for better yielding purpose is a difficult task due to vast parameters. These parameters differ from one farmer to another (depends upon their field). However, the climatic information collected in India at every 1sq.m area in different parts of the district are tabulated by Indian Meteorological Department. The huge such data sets can be used for predicting their influence on major crops of that particular district or place. There are different forecasting methodologies developed and evaluated by the researchers all over the world in the field of agriculture or associated sciences [2].

By analyzing the existing parameters such as soil or land productivity it is easy to predict the season for specific crop regarding its yielding effects [3] .Now days many prediction techniques are available for data analytics and that makes tremendous changes in agriculture sector. Next to prediction algorithm, the classification algorithms are also playing an important role for betterment in yield. Basically, the farmer knows the various plans for better yielding of crops but, in some cases it becomes extinct due to possible flood, so it is important to tangled up with the weather forecasts can help by letting the farmers realize much faster as to how to protect their living, by giving away predictions that take into account these extreme scenarios[4].

New information has to reach end-users very fast in order to use potential opportunities and achieve benefits. Information on seed, water, nutrients and plant protection is one of the main factors for successful farming [5]. The proposed system discuss about predicting yield production. Yield prediction is a very important in agriculture sector for the farmers, prediction will help to overcome situation like drought, rainfall etc. Some data mining techniques such as K-Means and Multiple Regression methods can be used to provide the solution for predicting yield production. This aims to data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities [6]. In India there are several ways to increase the economic growth in the field of agriculture. There are multiple ways to increase and improve the crop yield and the quality of the crops. The prediction method also produces better choice to the farmers by means of scheduling it, which additionally produce better outcome. Particularly, the machine learning (ML) techniques correlate more strong relationship between the farmers by opting the factors of crops and analyzing it. Thus, machine learning technique leads to better prediction.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

There are certain conditions that involved in machine learning method such as (i) the input/output characteristics are unknowns (ii) automatic specifications is required. While comparing with the basic statistic oriented methods, the machine learning methods has advantage in developing the overall model. This model will further determine the required data. These advantages make to evaluate the factors of the crops which better description. Basically, the machine learning is a division of artificial intelligence model [7]. The samples of input can be determined by training it and hence the metrics like mean square error, root mean square error, precision, recall, sensitivity specificity etc are taking impart.

The machine learning technique is enumerated with various types such as, including crop yield prediction through supervised, unsupervised and reinforcement learning methods. Classification, clustering, regression, prediction are some of the techniques involved to attain the intelligent system. Meteorological conditions, such as precipitation, temperature, soil conditions, topography and socio-economic factors are responsible for about 30% growth of the crops. Several works were proposed in the literature for predicting the yield of crop using expert systems, regression, , artificial neural network, data mining, support vector machine etc. The expert system such as fuzzy logic is based on logical rules to predict the yield. However, it requires wide interaction with the experts for obtaining the rules for prediction. In addition, these rules are based on certain set of input data [8].

II. LITERATURE SURVEY

Mahabadi et al. .[9] suggested a method using artificial neural network which consists of numerous neurons with hidden layer and front/back propagation unit. The author made some modifications in this unit in order to produce good yield. As a result, the rate is to be finding out by making a tremendous effect in hidden layer part. This effect produces more knowledge and resultant in overanalyzing of inbuilt learning models. For execution process the author opted root mean square error for better result.

Ranjeet et al.[10] used back propagation based artificial neural network for analyzing the yield of crops. This contains inbuilt neurons and hence the training is conducted with fault prediction for evaluating the environmental conditions, the factors such as temperature, type of fertilizer, area width are considers for better employable results. This work suggested a future work as collecting less loss-effective parameters, precise range of layers, picking the appropriate number of neurons by using optimization algorithm.

Meena et al. [11] proposed artificial neural networks by examining 24 factors and more datasets. These 24 factors are collected with respect to various months and hence the highest temperature is found out as average value. The supporting facto such as level of agriculture area, climatic condition, wind speed are opted for analyzing the crop quality.

Vats et al. [12] examined datasets such as soil, rain and previous yield. Now, the above mentioned datasets makes to cross check with the variety of available crops. Hence, they are further classify into vartious subdivisions using supervised machine learning method. This method is further coincide with linear regression model. During the classification process the categories such as low, medium and high are noted and hence the regression module showed proper crop yield. For supervised learning some of the techniques such as kernel neural network, support vector machine and least square support vector machine are used.

Vinciya et al., in [13] developed a technique namely multiple linear regression (MLR) with decision tree classification process. This tree classification is applied to 362 numbers of datasets. The datasets are planned to subdivide as organic and in organic. They are further used with comparison of various type of soil to predict the quality of soil. The extracted results are compared in terms of accuracy.

Shivnath et al., in [14] adopted back propagation based neural network to find out the training and testing data. Just like all the neural network, this also used its hidden layer for the prediction process. Moreover, the network is also designed with self-applicable version to evaluate the productivity of the soil. A s result, accuracy is calculated as high rate when comparing with existing methods but there is limitation in computational time and speed.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

Hong et al.,[15] two regression supervised machine learning methods are used: Support Vector Machine(SVM) and Relevance Vector Machine(RVM) to show effectiveness in soil quality prediction. a smart wireless device for sensing soil moisture and meteorological data. These methods makes better accuracy as 95% and error rate as 15%

Arik et al., [16] involves a check for Soil Fertility and Plant Nutrient by using back propagation algorithm. The results are accurate and enables improvement in soil properties. It performs better as compared to traditional methods. However, system is slow inefficient and not stable.

III. PROPOSED METHODOLOGY

In this section, we have concentrated on a methodology to predict the suitable crops for the particular soil. This is done by using Modified Multi-Layer Perceptron Neural Network (MMLPNN) algorithm. It contains 2 important steps such as prediction of potential oriented attributes with input and output. The training phase is done by using back propagation method.

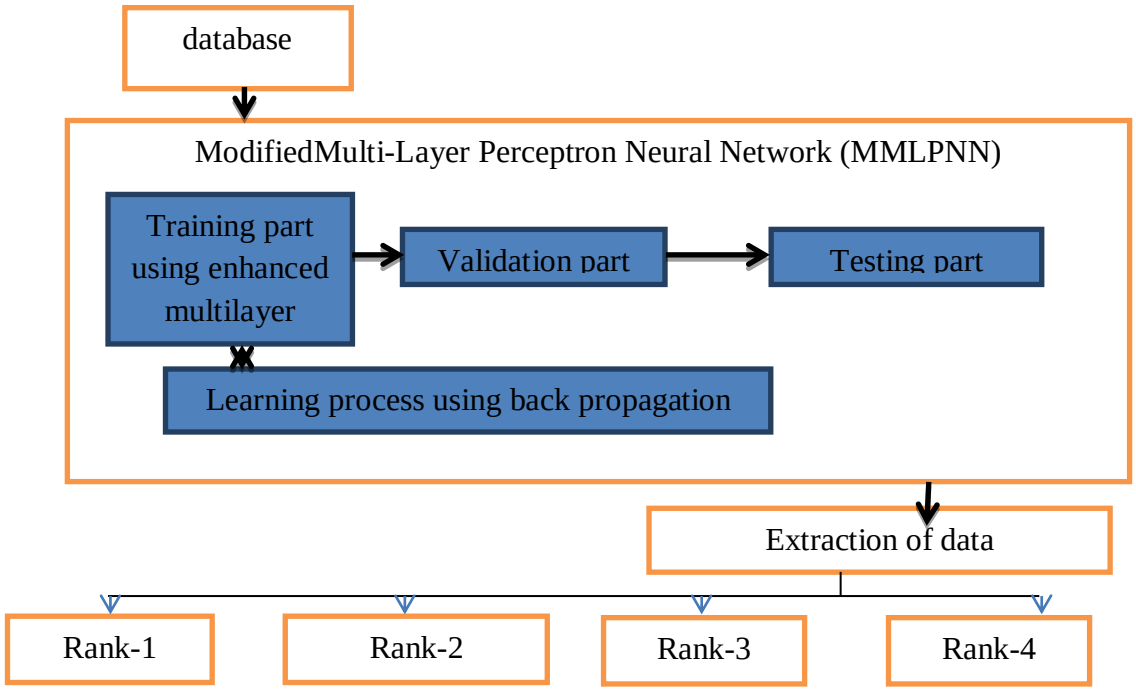

Figure-1 block diagram of proposed model

Data base

As mentioned above, the datasets are determined using basic attribute for th purpose of ranking various crops. Apart from that, some of the factors such as climate, rainfall, solar effects and temperature are also considered. In our dataset the weather data is obtained from database of National Meteorological Service (NMS) with the locality based content. The proposed methodology is utilized to furnish, similarize and connect the double database into single one. The potential attributes are opted based on existing available data. For better enhancement, the proposed method is hybrid with regression technique. As a result, a set of reusable subsets are consolidated. There are three important factors such as radiation; temperature and humidity are noticed for previous three stages of crops.

Attribute-1:Area for planning the crops (PA)-surface area for sowing and reaping the crops.

Attribute-2: water depth (WD) in centimeter-water used for irrigation purpose both in surface and underground way. This helps in development of crops in various stages.

Attribute-3: Radiation by solar (SR) in KWh/m^{2-}it indicate the average rate of radiation over 3 months.

Attribute-4: Rainfall (RF) in millimeter-Average rate of rating over last cultivation period.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

Attribute-5: Maximum temperature (°C) A-It indicates the temperature rate on daily basis.
Attribute-6: Average temperature (°C) Average of daily mean temperatures registered in the last three crop growing stages
Attribute-7: Minimum temperature (°C) Average of daily minimal temperatures registered in the last three crop growth stages
Attribute-8: Season-duration cultivar Identifies the kind-duration cultivar of the crop (1 = short, 2 = medium, 3 = long). Duration time (in days) is different depending on each crop type.

Training process:

Initially, the input layer of Multi-Layer Perceptron Neural Network (MMLPNN) is subdivided into Xi signals within the hidden layer. Now, each neuron 'j' is coinciding with input signal 'i' along with the weighted rate of Wij from the appropriate input layer. As a result, the output Yj is produced with the summation function (F) as follows.

$$Y(i) = F\left(\sum_{k=0}^{n} W(ji)X(i)\right)$$

The working of Multi-Layer Perceptron NeuralNetwork is summarized in steps as mentioned below:
1) Input data is provided to input layer for processing, which produces a predicted output.
2) The predicted output is subtracted from actual output and error value is calculated.
3) Apply stochastic based nelder mead algorithm (SNMA) for minimizing the complexity with back propagation method. It is achieved by evaluating the subsets of each training data with specified iterations. This will end up with reducing the overall error by achieving the optimum level. Here, every training data acquires more number of iteration with the adjustable weights on epochs.
4) For weights adjusting it starts from weights between output layer nodes and last hidden layer nodes and works backwards through network.
5) When back propagation is finished, the forwarding process starts again.
6) The process is repeated until the error between predicted and actual output is minimized.

Prediction:

In this prediction system the database is used to fed the input to neural network. The network is having three layers and feed forward neural network modelThe input layer consists of 22 neurons and hence the 22 datasets are loaded on it. Now, the hidden layer is activated with ranking concept with the minimization of error rate.

Types of crops for ranking
Rabi crops
Zaid crops
Kharif crops
Jowar crops
Cash crops
Fibre crops
Wheat

A combinatorial procedure to perform a complete enumeration of all the subsets (X1,X2,X3….XN) is presented in this paper. The procedure starts with a potential set of attributes B=(b1,b2,b3….bn) , such that each XI is a subset of A. Each subset is evaluated using the training dataset, which is divided in two datasets. The majority of samples are used to build the models, while the most recent ones are applied for performance measurement. In context, if the year range of historical data is available for training, the [A,B-1] range is really used for training, and data from year is reserved for validation.. Each validation result and the related attribute subset are registered in a sorted list according to these metrics. Ties are solved in the following order: RRSE (lower), (higher), and RMAE (lower). At the end of the process, the subset at the top is taken as the best. In testing stage, the algorithm is applied to the union of the training and the testing datasets, obtaining a rank of attribute subsets. In this last case, the subset at the rank's top is named the optimal attribute subset (OAS). Evidently, this last rank cannot be available in practice, because testing dataset represents unseen samples from the future. However, the rank of attribute combinations that originated the OAS can be used to define a new performance metric, which should be used only for evaluation purposes. Let X be an attribute subset

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

and D the number of combinations that separates the OAS results from the X subset results. Then it can be used as a performance measure of X. We called measure D the "distance to the optimal attribute subset.

Performance analysis

We use three of the most common metrics: the root relative square error (RRSE), correlation factor (R), and the relative mean absolute error (RMAE). RRSE compares the model prediction against the mean, which is frequently used to supply the crop yield value. An RRSE less than 100% indicates a prediction that is better than the average value. Correlation factor () measures the linear relationship between regression model predictions and the real values. Mean absolute error (MAE) is the average of estimation differences (in physical units). This metric is expressed as a percentage relative to the mean yield, being called RMAE instead of MAE. Equation (1) shows how these metrics are calculated, where y is the real yield value, y* represents the yield estimation, is the number of sample, is the average of the real yield values, and is the average of predictions:

$$\text{RRSE } (\%) = \frac{\sqrt{(y-y*)\text{^}2}}{\sqrt{(y-y*)\text{^}2}} * 100$$

$$\text{R} = \frac{\sum_{I=1}^{N}(y-y*).(y*-y\text{^})}{\sqrt{(y-yi)^2}.\sqrt{(y1-y*)\text{^}2}}$$

$$\text{RMAE}(\%) = \frac{(yi-yi*)}{(n.y*)} * 100$$

parameters	Modified Multi-Layer Perceptron Neural Network (MMLPNN)	Artificial Neural Network	Recurrent neural network
Sensitivity	91%	72%	81%
Specificity	89%	79%	76%
accuracy	99%	61%	73%

The statistical measures that can be considered are sensitivity, specificity, and accuracy

$$\text{Sensitivity} = \frac{TP}{TP+FN} * 100$$

$$\text{Specificity} = \frac{TN}{TN+FP} * 100$$

$$\text{positive prediction value (PPV)} = \frac{TP}{TP+FP} * 100$$

$$\text{negative prediction value (NPV)}, = \frac{TN}{TP+FN} * 100$$

Given two classes, we can say in terms of positive records versus negative records. True positives refer to the positive records that have been correctly labeled by the classifier, while true negatives are the negative records that have been correctly labeled by the classifier. False positives are the negative records that have been incorrectly labeled.

Similarly, accuracy of a classifier for a given test dataset is indicated by the percentage of test dataset records that are correctly classified by the classifier. The sensitivity and specificity are alternatives to the accuracy measure used to evaluate the classifier's performance.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

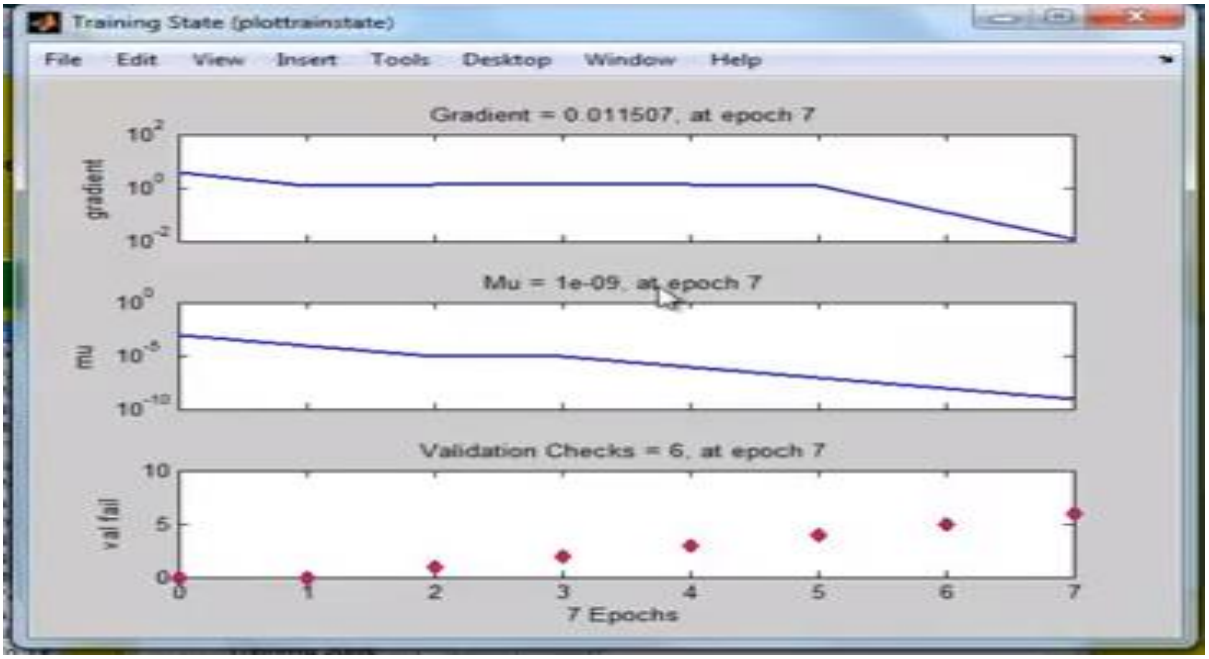

Figure-2 Performance of the system for 7 epochs

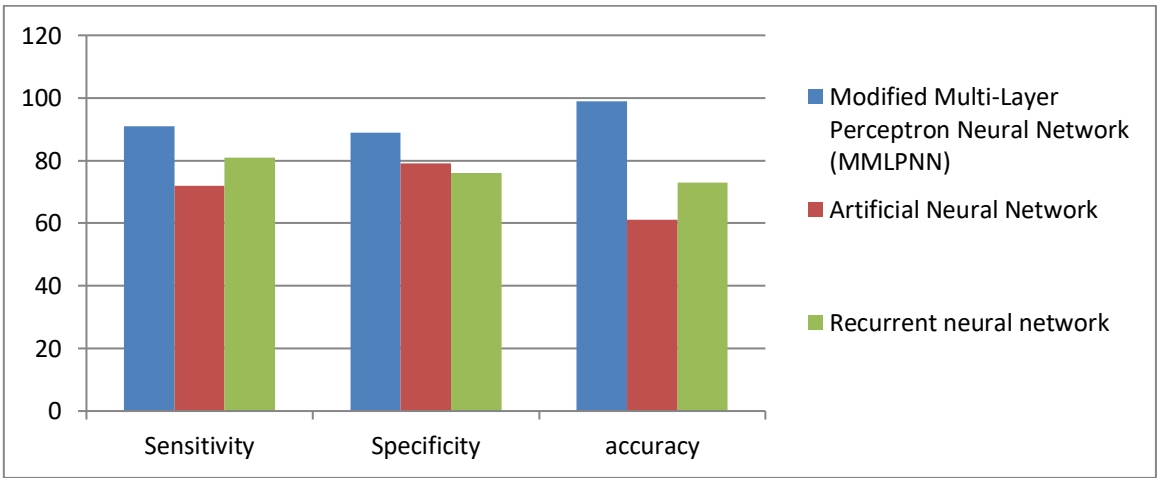

Figure-2 comparison of statistical parameters

IV. CONCLUSION

Predictive analysis in agriculture is crop yield monitoring concept. Implementation of this concept should help farmer to produce higher yield. With this concept, input factors can be minimized and output can be maximized in precise way. In order to increase production, the system manages crop diseases also. System can find required amount of land, water, fertilizer, pesticides to maximize the crop production. This will help the farmers to plough exact resources for production and minimize the cost in proper manner. This technique significantly reduces cost and chemical wages. So that farmers obtain a return on their investment by saving on fertilizer cost. Another benefit of this concept is adjusting with environmental factors. Based on different environmental structures farmer can plough respected crop on respected field at right time increases the productivity to larger manner. Therefore precision agriculture seeks important role in economy field as well as to achieve environmental goal.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 6, Issue 5, May 2019

REFERENCES

1. M.Shashi, Y. a. (2009). Atmospheric Temperature Prediction using Support Vector Machines. International Journal of Computer Theory and Engineering, 1 (1).
2. Miss.Snehal, S. D. (2014). Agricultural Crop Yield Prediction Using Artificial. International Journal of Innovative Research in Electrical,Electronic, 1 (1).
3. Obua, W. O. (2011). Machine Learning Classification Technique for Famine Prediction. Proceedings of the World Congress on Engineering, 2.
4. SnehalS.Dahikar, Dr.SandeepV.Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering, 2(1),2014, 683-686
5. D Ramesh and B Vishnu Vardhan,"Analysis Of Crop Yield Prediction Using Data MiningTechniques", International Journal of Research in Engineering and Technology, 2015,4(1)
6. Mahabadi , tahisurei " Use of the Intelligent Models to Predict the Rice Potential Production, International Academic Journal of Innovative Research, Vol. 2, No. 10, 2015, pp. 20-31.
7. Ranjeet "An Artificial Neural Network for Predicting Crops Yield in Nepal, Asian Federation for Information Technology in Agriculture, pp. 371-381
8. Vats ,Ankalaki, S., Chandra, N., Majumdar, J., "Applying Data Mining Approach and Regression Model to Forecast Annual Yield of Major Crops in Different District of Karnataka", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, SpecialIssue 2, pp.25-29, 2016
9. P. Vinciya, Dr. A. Valarmathi, "Agriculture Analysis for Next Generation High Tech Farming in Data Mining" IJARCSSE,vol. 6, Issue 5, 2016.
10. ShivnathGhosh,SantanuKoley, "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks" IJRITCC, vol. 2, Issue 2,292-297,2014.
11. ZhihaoHong,Z. Kalbarczyk,R. K. Iyer, "A Data-Driven Approach to Soil Moisture Collection and Prediction" IEEE Xplore,vol. 2, Issue 2,292-297,2016
12. SabriArik, Tingwen Huang, Weng Kin Lai, QingshanLiu , "Soil Property Prediction: An Extreme Learning Machine Approach" Springer, vol. 3, Issue 4,666-680,2015.