IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

# Prediction of Soil Quality using Machine Learning Approach

**Ramya R[1], Ranjitha D[2], Revathy T[3], P R Vijeth[4], Ranjitha U N[5*]**

[1,2,3,4,5]School of Computing and Information Technology, REVA University, Bengaluru, Karnataka, India

[*]*Corresponding Author: ranjitha.un@reva.edu.in, Tel.: +91-9900499044*

DOI: https://doi.org/10.26438/ijcse/v7si14.279283 | Available online at: www.ijcseonline.org

***Abstract***—our idea is to develop a machine learning model which is capable of predicting the quality of soil. Our idea focuses on Agriculture domain. Agriculture is the key to the economy and infrastructure of India. It plays a significant and most strategic role in the progress and financial growth of the nation. As the technology is rapidly advancing, extending it to agricultural domain yields most needed and promising results in achieving precision agriculture. The model we have designed is a works towards achieving it. The model that analyzes the quality of soil thereby predicting the yield of the crop by considering various parameters. Crop yield prediction provides information for decision makers to maximize the crop productivity.Manually testing the quality of soil regularly is a complex task, so there is a need for automating the process that we are currently following, through an ML (Machine Learning) Model. Machine learning approach offers new contingency in the field of agriculture which is very much useful in soil dataset analysis and visualization of various parameters related to soil which would also help in decision making. It is crucial to design and implement a well-planned management system for monitoring various nutrients level by means of soil analysis procedure. In our model, various soil data sample from various regions are classified based on primary and secondary properties.

***Keywords***—Machine Learning, Image Dataset, Soil Parameters, Image Processing, Supervised Learning, SVM Image Classifier

## I. INTRODUCTION

Agriculture is termed as the backbone of our Indian economy. While taking a glance at the difficulties that are being faced by our farmers, it has been reportedly seen that there is huge growth in suicide rate of farmers over the years. One of the major reason causing such a miserable condition to our farmers are increased debt and poor yield. In some cases, farmers are not aware about the crop which suits their soil quality, soil nutrients and soil composition. Soil is an important agricultural resource and maintaining the quality of soil leads to good agricultural productivity. Soil supports the growth of plant and represents the living reservoirs that represent the flow of water, micronutrients, and energy through a biological ecosystem. Soil testing is important for increasing agricultural production. Testing of soil for its quality is the only way for farmers to know how to treat their soil in a better way that can give them expected and good yield. But the traditional way i.e. hardware-based testing of the soil is a very complicated task and also it fails to produce precise results. The solution which has been suggested would overcome these drawbacks and predicts the quality of soil by analyzing unstructured data. Machine learning uses analytical methods to make the computers learn on their own and progressively improve performance of a particular task using the given data. Hence it can be used to resolve some of the current issues related to agricultural domain. This software implementation does not require much maintenance. The accuracy level is high when compared to hardware based systems, because parameters like composition of soil, type of soil, pH, and weather conditions are all taken into consideration. The work proposed encourages the farmers to check the soil quality of their farm land based on machine learning approach. Thus the proposed system emphasizeson checking the soil quality and suggestssome of the cropsthat give high yield according to their soil type and related parameters.


*Fig 1: Dependency of Agricultural Sustainability*

 

Improved agricultural practices are important to tackle the problems of agriculture in terms of productivity. These challenges in the field of agriculture can be addressed at the basic level by predicting the quality of soil through Machine Learning approach and hence increasing the crop production.By using ML and Image Processing approaches, wehave created a model that is capable of predicting the quality of soil automatically without using any manual testing methods. A remarkable step which we have adopted in this work is considering the images of soil as our dataset instead of using for normal datasets like excel, csv etc. Also, the algorithm used, SVM image classifier, gives maximum accuracy when compared to other supervised learning algorithms.

## II. RELATED WORK

### 1. Agriculture Analysis for Next Generation High Tech Farming in Data Mining[1]

This paper presents a brief analysis of crop yield prediction using Multiple Linear Regression (MLR) technique for the selected region. This project mainly focused on analysing organic farming and inorganic farming, time period of cultivation of the plant, profit and loss of the data and analyse the real estate business land in a specific area. The purpose is to estimate difference in efficiency and prediction between organic and inorganic farming. This work aims at finding suitable data models that achieve a high accuracy and a high generality in terms of yield prediction capabilities.It provides the future information of the agriculture status at the agriculture department.Data mining techniques are mainly divided in two groups, classification and clustering techniques.The meta-analysis resulted in the inclusion of 362 paired sets of organic–conventional yield data in the database.

### 2. Machine learning for soil fertility and plant nutrient management using back propagation [2]

The objective of this paper is to analyse the main soil properties such as organic matter, vital plant nutrients and the micronutrients that affects the growth of crops and find out the suitable relationship percentage of the properties by using Supervised Learning Technique and Back Propagation Neural Network. BPN will identify and predict the correct correlation percentage among the pre-defined properties. This process is divided into three parts, first sampling, second Back Propagation Algorithm and third weight updating rule. The representation of the Back Propagation Neural network model will be checked using a test data set. To conclude, the result of this study shows that training is very important to increase the accuracy of one region.

### 3. Prediction of crop yield using machine learning[3]

The objective here is to help farmers to check the soil quality depending on the analysis done based on data mining approach.Having a software approach, it does not allow maintenance factor to be reviewed much. Also the accuracy level would be high as compared to hardware based solutions, because components like composition of soil, type of soil, pH value, and weather conditions all come into picture during the prediction process.It can be achieved by using unsupervised and supervised learning algorithms. Dataset will be then trained by learning systems. It differentiates the precision obtained by various network learning techniques and the most accurate result will be delivered to the end user.

### 4. Soil characterisation based on digital image analysis [4]

The traditional laboratory approach has some drawbacks such as manual involvement, time consumption, chances of committing errors and uncertain prediction. This research work is undertaken to develop a correlation between the feature of soil image namely, Fractal Dimension(FD) and physical properties of soil material. The digital image database is prepared for the collected soil and physical properties(Y) are determined. Digital image analysis is adopted to estimate the image feature namely fractal dimension (X). Correlation is developed between Y and X by fitting appropriate polynomial equations using regression models. The final results of this research contributes to make estimation of the soil's physical properties and also emphasizes on the importance of fractal dimensionfor estimating physical soil properties with minimum human error.

### 5. Machine learning based prediction of soil total nitrogen,organic carbon and moisture content by using VIS-NIR spectroscopy [5]

It's widely known that visible and near infrared spectroscopy has the potential of estimating the total nitrogen content, organic carbon and moisture content. Improving the accuracy requires advanced modelling techniquesto compare the predictive performance of two linear multivariate and two machine learning methods. The two multivariate methods are Principal Component Regression (PCR) and Partial Least Squares (PLSR). Machine learning methods include Least Square Support Vector Machines(LS-SVM) and Cubist. The results indicate that machine learning methods are capable of tackling non-linear problems in the dataset. LS-SVM provided the best prediction for moisture content.

## III. METHODOLOGY

We have built an ML Model that is capable of predicting the soil quality and suggesting suitable cropsin that particular soil based on a set of predefined Primary and secondary parameters. Primary parameters includesoil granularity, Water holding Capacity, Organic carbon, soil depth that affect the quality of soil adversely and secondary parameters

 

include pH, nutrients that have less influence on the quality of soil.

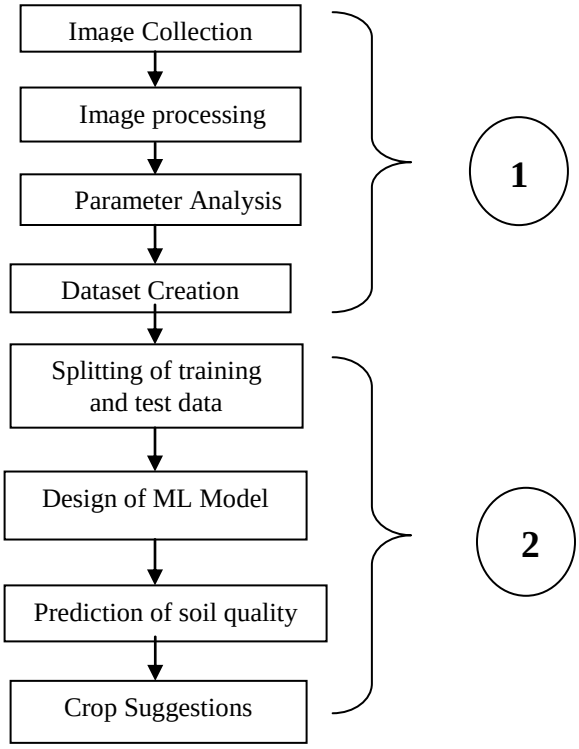


*Fig 2: Block diagram of the proposed system*

A soil image is given as input which is considered as the test image, this image is compared with the set of predefined images which are used as training data for the ML Model that we are designing and displays the various attributes like PH, crop suggestion, information about nutrients.

Step 1 of Fig 2: This step includes collection of images from various places, processing those images, analysing various parameters like PH, soil depth, organic carbon, water holding capacity etc. After analysis of various parameters the next step is to create a dataset with all the above processed images along with their parameters.

Step 2 of Fig 2: The above created dataset is split into training data and test data. Around 75% of the data from the dataset is considered as the training data and the rest 25% is considered as test data. An image from the test data is given as input and compared with the other images of the training data and displays the parameters as output.

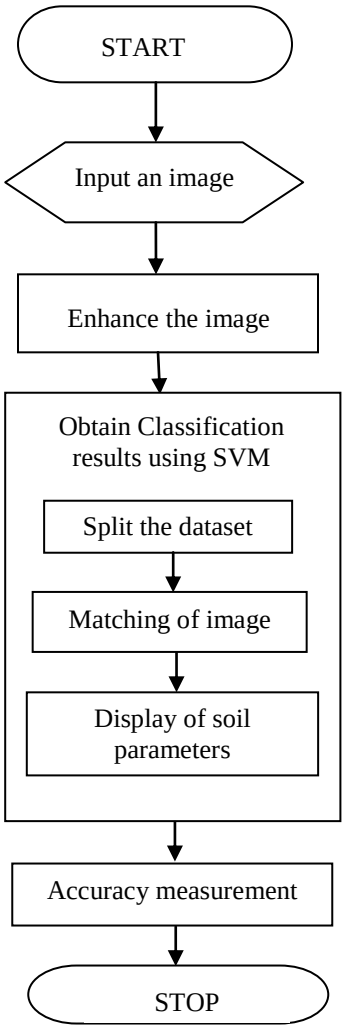


*Fig 3: Flowchart of the proposed system*

## IV. RESULTS AND DISCUSSION

An image is given as input data from which a characterised output which determines the quality of the given soil is obtained based on the primary and secondary attributes.The training example focuses on the following attributesNutrient Deficiency, Granularity, Water Holding Capacity, Soil Depth and Electrical Conductivity.

The algorithms which were taken into consideration during this work are elaborated in the following section along with their functionalities.

1. The Naive Bayes Classifier is a combination of classification algorithms based on Bayes' Probability Theorem. It is a group of algorithms that have a common principle. This algorithm gave an accuracy of about 70-72% which was too less and hence we were not able to classify the data as expected.

 

2. Decision Tree Algorithm is a supervised learning algorithm that can be used for solving regression and classification problems. Using this, we can create a training model which may predict the class or value of target variables by learning decision rules inferred from training data. This algorithm gave an accuracy of about 78-80% which was not sufficient to classify the data successfully.
3. The Support Vector Machines (SVM)are supervised learning models learning algorithms that analyse data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other. This algorithm gave an accuracy of about 95-98% when the dataset was subjected to testing.

Hence, among the above referred algorithms we have chosen Support Vector Machine to carry out our model as it gave the maximum accuracy on all tools.

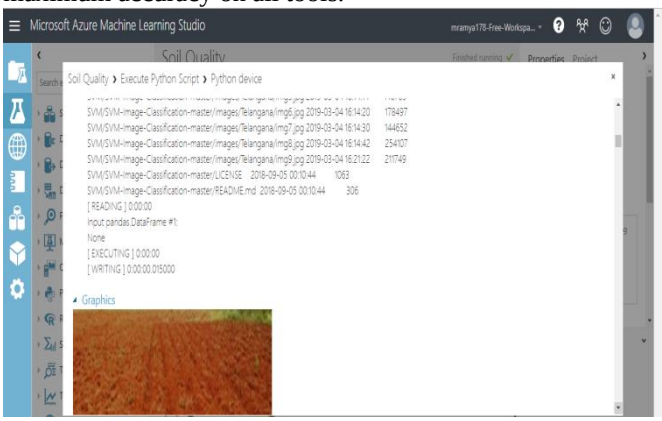


*Fig 4: Output using SVM image classifier*

The tools that we worked on during this study are elaborated in the following section along with their functionalities.

1. Waikato Environment for Knowledge Analysis (WEKA) is a machine learning software written in Javathat contains a wide range of visualization tools and algorithms for data analysis and predictive modelling. This tool supports only csv files and excels sheets to be taken as input. Since our dataset contains images which have to be given as input, this tool could not support this study.
2. Rapid Mineris a data science platform that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics.It is used in machine learning process for data preparation, results visualization, model validation and optimization. When our dataset containing images was given as an input in this tool, the images were being converted intoND array which was very difficult to visualize and extracted parameters of the soil were not as per the requirements. Also, the execution time of this tool is too long.
3. Orangeis an open-sourcedata mining tool with visual programming toolkit. It features a visual programmingand interactive data visualizationfor explorative data analysis, and can also be used as a Python library. While using this tool we were not able to implement many of the required functionalities of our study.
4. Microsoft Azure ML Studio is a collaborative, drag-and-drop tool that can be used to build, test and deploy predictive analytics solution on our data. Using this tool we were successfully able to get the expected results. A zip file containing the images dataset, SVM image classifier code in python was imported onto this tool. The code was executed and after resolving the errors, the results obtained were soil images of good quality along with their URLs. Since we were able to display only soils of good quality and not the parameters which helped in predicting the quality of soil, we were made to proceed with image processing using MATLAB.
5. MATLAB 2013a: MATLAB (Matrix Laboratory) is a multi-paradigm numerical computing environment and proprietary programming language developed by MathWorks. It allows us to do different operations like matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces and image processing.

After working on all the above mentioned tools, we found out that MATLAB satisfied almost all the requirements that we needed in our study including classification of soil images, image processing for better clarity of images and extraction of soil parameters.

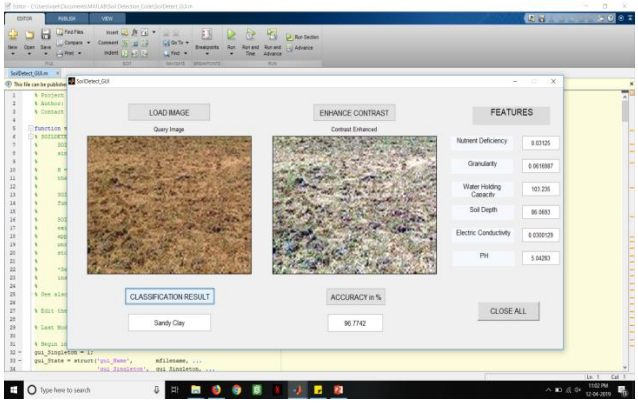


*Fig 5: Output using MATLAB*

The final result is that when an image of soil is loaded as input or query image in the GUI which was created using MATLAB, it is subjected to image processing through which we can get an enhanced image of the soil along with the list

 

of values that determine the quality of soil and its accuracy measurement. It also suggests the crops that can be grown in that particular soil to get maximum yield. An accuracy of about 95-97% was achieved when the input data was subjected to testing. Hence we can conclude that by using Image Processing along with Machine learning, we can predict the quality of soil in a better way to get accurate results.

## V. CONCLUSION AND FUTURE SCOPE

Our project stands unique in front of other projects related to our topic as we are successful in predicting the quality of soil by considering the images of soil along with many different parameters that are remarkable in measuring the quality of soil. Literature survey revealed the fact that the number of parameters considered were too less and were readily available as datasets. Here, we have created our own dataset with more number of parameters by studying the surveys on soils from different regions conducted by various government organizations like Central Ground Water Board, Indian Institute of Plantation Management etc.

Also, we have included SVM image classifier algorithm along with few functions of MATLAB, which is one of the interesting feature that enhances the uniqueness of our project.

Hence, by integrating Machine Learning and Image Processing, we can create a model that is capable of predicting the quality of soil automatically without using any manual testing methods.

Applications include:

- Automation in achieving precision Agriculture.
- The results obtained might help to designvarious agricultural strategies from The Governmentin order to manage the soil fertility degradation,crop productivity and usage of fertilizers.
- This research may also help us to avoid expensive chemical measurements of OC, $N_2O$, $P_2O_5$ and $K_2O$ and reduce bio-magnification in the food chain
- Saves time of professional technicians in designing and developing expensive chemical analysis methods.

This work will encourage the government to make decisions about improving soil quality and annual crop production.

## REFERENCES


[1] P. Vinciya, Dr. A. Valarmathi, "*Agriculture Analysis for Next Generation High Tech Farming in Data Mining*", International Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE), Volume 6, Issue 5, 2016.

[2] Shivnath Ghosh, Santanu Koley, "*Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks*", International Journal on Recent and Innovation Trends in Computing and Communication(IJRITCC), Volume 2 Issue 2, 2014.

[3] Priya R L, "*Prediction of crop yield using machine learning*", International Research Journal of Engineering and Technology (IRJET), Volume 5, Issue 2, 2018.

[4] Karisiddappa, Ramegowda, "*Soil characterisation based on digital image analysis*", In the proceedings of 2010 Indian Geotechnical Conference, India.

[5] A. Morellos, "*Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy*", Biosystems Engineering, Volume 152, 2016.

[6] Vrushal Milan Dolas, *"A Novel Approach for Classification of Soil and Crop Prediction",* International Journal of Computer Science and Mobile Computing, Volume 7, Issue 3, 2018.

[7] Supriya D M, *"Analysis of Soil Behavior and Prediction of Crop Yield using Data Mining Approach",* International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 5, 2017.

[8] Rushika Ghadge, *"Prediction of Crop Yield using Machine Learning"*, International Research Journal of Engineering and Technology, Volume 5, Issue 2, 2018.

[9] Vaneesbeer Singh,Abid Sarwar, *"Analysis of soil and prediction of crop yield (Rice) using Machine Learning approach"*, IJARCSE, Volume 5, Issue 8,2017.

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


 