International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Soil classification and crop suggestion based on deep learning

T. Thanish kumar, B. Akhileshwar Reddy, B. Bharath Chandra, S. B. Charan Raju, B. Deepak kumar
School of Engineering, Mallareddy University, Hyderabad – 500043, Telangana, India;
kumarthanish@gmail.com, bommineniakhileshwarreddy@gmail.com , bc2990772@gmail.com, Charanraju023@gmail.com, DeepakKumarbachu@gmail.com

Abstract

Agriculture is the major source for living for the people of India. Agriculture research is the major source of economy for the country. Soil is an important key factor for agriculture .There are several soil varieties in India. In order to predict the type of crop that can be cultivated in that particular soil type we need to understand the features and characteristics of the soil type. Machine learning techniques provides a flexible way in this case. Classifying the soil according to the soil nutrients is much beneficial or the famers to predict which crop can be cultivated in a particular soil type. Data mining and machine learning is still an emerging technique in the field of agriculture and horticulture. In this paper we have proposed a method for classifying the soil according to the macro nutrients and micro nutrients and predicting the type of crop that can be cultivate in that particular soil type. Several type of machine learning algorithms are used such as K-Nearest Neighbor (k-NN), Bagged tree, Support vector machine(SVM) and logistic regression.

Keywords: Machine learning, agriculture, soil, classification, nutrients, chemical feature, accuracy.

1. Introduction

Data mining has been used for analyzing large data sets and establish classification and patterns in the datasets. The techniques are used to elicit significant knowledge that can be easily predictable by individuals. Data mining is a challenging technology in the field of agriculture. Nowadays data mining has been used in the field of agriculture for soil classification, wasteland management, crop and production. The main reason for agriculture is to grow crops. Crop cultivation depends on the nature and the nutrients of the soil increasing the cultivation of land which brings a loss of supplements present in the soil. In the crop cultivation soil plays an important role. It is important for plants, animals rocks and living organisms. All of these helps in managing the fertility of the soil. A soil test is carried out to identify the nutrients content, composition and other components contained in the soil.Soil tests are mainly conducted to measure the fertility and other defencies present in the soil so that suitable measure can be taken to resolve it.

Machine learning is a field of computer science where new developments evolve at recent times and also helps in automating the evaluation and processing pest management. In assessed the association rules of affiliation methods in DM and applied into the soil science to anticipate the significant connections and gave association rules to different soil types in agriculture. The agriculture factors such as rain, weather, soil type, pesticides and fertilizers are the

main responsible to increase the done by the mankind ,thus by reducing the burden on human power. Machine learning is the field of Artificial Intelligence by the dint of which computers can be taught without explicit programming. In simple terms, the meaning of machine learning is the basic algorithms can provide information about a dataset without writing code to solve this program manually . Instead of writing code you provide data or the basic algorithm and it forms its own conclusions based on this data . In machine learning agriculture, the methods are derived from learning process . Those methodologies need to learn through the experiences to perform a particular task. Once the learning is completed then the model can then be used to make an assumption to classify and to test be used to make an data. The data is achieved after gaining the experience of the training process.

Classification is the main problem in data mining. Classification is a data mining technique based on machine learning which is used to categorize the data item in a dataset into a set of predefined classes. It helps in finding the diversity between the objects and concepts. It also provides necessary information for which research can be done in a systematic manner.

2. Related Work

In a research carried out by Zaminur Rahman, a comparative study of several machine learning techniques has been carried out. They have carried out the classification using the data of Bangladesh. Considered the six district soil data and used the geographical features for classification. They have used k Nearest Neighbor, Bagged tree and SVM finally compared the results of three algorithms and brought out a model for classifying the soil types and the suitable crop that can be cultivated in that particular soil type. Among the used three algorithms SVM has obtained the average accuracy. In a research carried out by Leisa J. Armstrong a comparative study of data mining algorithms. They have used a large dataset extracted from the Australian Department of Agriculture and Food(AGRIC) to conduct the research.

In an approach carried out by Jay Gholap carried out a modal to classify the soil based on fertility. The dataset was collected from the soil testing laboratories of Pune District. They have used WEKA tool for developing an automated system.

Chiranjeevi M. N carried out a research for classifying the soil types so that it can be useful for the farmers for analyzing the type o soil and the crop that can be cultivated so that there will a good yield and profit. They have considered the data mining algorithms for classifying the soil. They have used algorithms such as J48 decision tree classifier and Naïve bayes classifier among these two algorithms Naïve bayes has obtained the maximum accuracy of 98%.

V. Rajeshwari and K. Arunesh, Analyzing Soil Data using Data Mining Classification techniques, Vol 9(19),May 2016.

Jay Gholap , Anurag Ingole , Jayesh Gohil, Shailesh Gargade, Vahida Attar (2013), Soil data analysis using classification techniques and soil attribute prediction,.

Sk Al Zaminur Rahman, Kaushik Chandra Mitra ,S.M. Mohidul Islam(2018),Soil classification using Machine Learning Methods and Crop Suggestion based on Soil Series.

L.Armstrong , D.Diepevven & R. Maddern(2004),The Application of Data Mining Techniques to categorize agricultural soil profiles.

Ramesh Vamanan, K.Kumar (2008),Classification of Agricultural Land Soils A Data Mining Approach.

Chandrakar PK , Kumar S, Mukherjee D(2011), Applying classification techniques in Data Mining in agricultural land soil.

Campus Valls G , Gomez Chova L , Calpe Maravilla J, Soria

Olivas E, Martin Guerreo JD, Moreno J(2003) Support vector machines for crop classification using hyperspectral data.

Bhuyar V(2014), Comparative analysis of classification techniques on soil data to predict fertility rate for Auranagbad District .

T .Mathavi Parvathi ,Automated soil testing process using combined mining processManonmaniam Sundaranar University.

The methodology involves the following steps:

- Dataset collection
- Pre-processing
- Classification
- Prediction
- Result

A. Data collection

Most of the research papers carried out the model using the chemical parameters , water content, electrical conductivity, organic content and the fertility. The values of these are taken as inputs for the algorithm.

B. Pre-processing

For a successful completion of a model a huge set of data is required. The data that is collected from real world might be in raw format .It may contain some missing values, inconsistent and noisy values. In this step such redundant values should be filtered. The data is made normalized.

C. Classification

It is one of the data mining .This is used to analyze the data and allocates it into a separate class. In pre processing step a prototype is developed. In classification the removed prototypical is tested against the pre defined dataset. That is to quantify the prototypical trained performance and accuracy. D. Prediction. The presentation of classification algorithm associated based on accuracy and performance analysis and will provide a suggestion for the farmers to cultivate in a particular soil type. The final result gives the suggestion of crops.

3. Proposed Methodology

The System architecture of the proosed model is shown in fig. 1

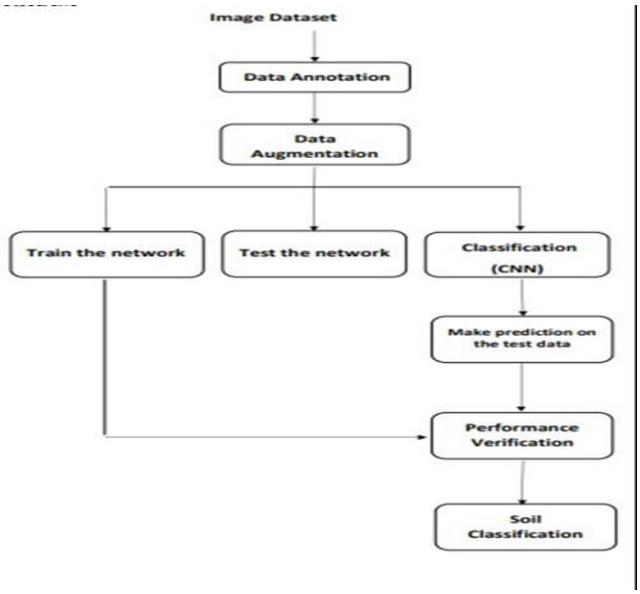


Fig. 1: Proposed System Architecture

The method involves two phases: training phase and testing phase. Two datasets are used: Soil dataset and crop dataset. Soil dataset contains class labeled chemical features of soil. Table I shows the details of the 12 chemical attributes of soil, used in our method.

TABLE I. -- CHEMICAL ATTRIBUTES

Attribute	Details
pH	pH value of soil
Salinity	Ds/meter
Organic Matter %	Percentage
Potassium	Mili equivalent/100 gram soil
Sulphur	Microgram/per gram soil
Zinc	Microgram/per gram soil
Boron	Microgram/per gram soil
Calcium	Mili equivalent/100 gram soil
Magnesium	Mili equivalent/100 gram soil
Copper	Microgram/per gram soil
Iron	Microgram/per gram soil
Manganese	Microgram/per gram soil

Soil series and land type combinely reprsents the soil class in the database. The machine learning methods are used to find the soil class (i.e. soil series and land type). Three diffrent methods are used: weighted K-NN, Gaussian Kernel based SVM, and Bagged Tree.

A. Weighted K-NN

A refinement of the *k*-NN classification algorithm is to weigh the contribution of each of the *k* neighbors according to their distance to the query point x_q , giving greater weight w_i to closer neighbors. It is given by

$$F(x_q) = \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i}$$

Where the weight is,

$$w_i \equiv \frac{1}{d(x_q, x_i)^2}$$

In case of Xq Exactly matches one of Xi so that the denominator becomes zero, we assign F(Xq) equals F(Xi) inthis case , It make no sence to use all training examples not just *k* if weighting is used, the algorithm then becomes a global one. The only disadvantage is that the algorithm will run more slowly.

B. SVM

SVM is a supervised machine learning algorithm which works based on the concept of decision planes that defines decision boundaries. A decision boundary separates the objects of one class from the object of another class [11]. Support vectors are the data points which are nearest to the hyper-plane. Kernel function is used to separate non-linear data by transforming input to a higher dimensional space. Gaussian radial basis function kernel is used in our proposedmethod.

$$K(X_i, X_j) = e^{-\|X_i, X_j\|^2 / 2\sigma^2}$$

Where K(Xi,Xj) = Feature vectors in input space ,||Xi,Xj||^2 = High Dimensional space of X and Y co-ordinate, and sigma is a free parameter.

C. Bagged Tree

We have used a bagged decision tree ensemble classifier (consisting of 30 trees). Bagging generates a set of models each trained on a random sampling of datad (Bootstrap resampling). The predictions from those models are aggregated to produce the final prediction using average(shown in fig.2).

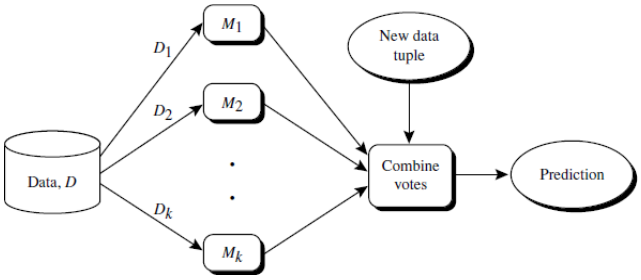


Fig. 2. Ensemble tree

TABLE II . SOIL CLASSES

Soil Class (Soil Series name with Land Type)	Class Label
Gopalpur High Land	1
Isshwardi Mid High Land	2
Ghior Mid High Land	3
Ghior Mid Low Land	4
Bajoya Mid High Land	5
Barisal Mid High Land	6
Barisal Mid Low Land	7
Harta Mid Low Land	8
Jhalokathi Mid High Land	9
Dumuria Mid High Land	10
Komolkathi Mid High Land	11

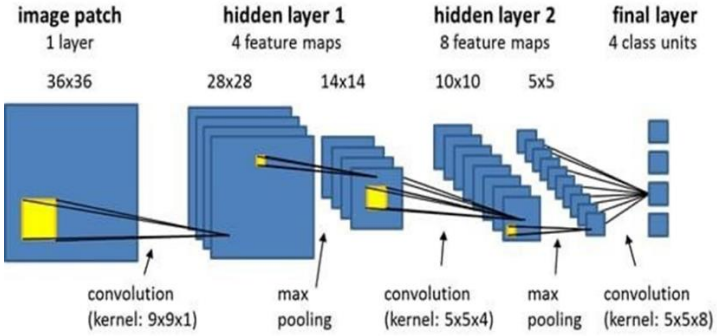
Two-third of the samples are used to train the model(s), rest one-third of the samples are used to test the models. For correctly classified samples, crop is suggeted for the corresponding map unit of corresponding upazilla. These two geogrphical features and suggested crop-list makes the crop dataset. The total map units of six upazillas of Khulna district, Bangladesh is shown in table III.

TABLE III. TOTAL MAP UNITS

Upazilla Name	Upazilla Code	Total Map Units
Rupsha	100	7
Dighalia	200	10
Fultola	300	12
Koyra	400	4
Dakop	500	8
Terokhada	600	9

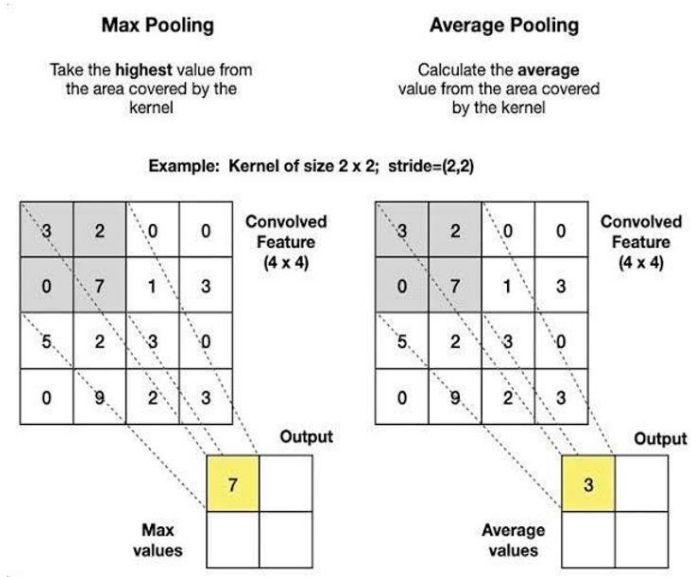
CONVOLUTION LAYER

The 4x4x3 image's dimensions should be decreased in the first stage. We define a filter with a dimension of 2x2 for each color to achieve this. Additionally, we desire a step length of 1, meaning that the filter should advance exactly one pixel after each computation step. The image's details will be kept while the dimension is not drastically reduced. A 3x3 matrix will be produced by our convolutional layer if we replace a 4x4 matrix with a 2x2 matrix and move one column or one row in each step. The figure illustrates how the scalar product of the two 2x2 matrices is used to calculate the matrix's component values.



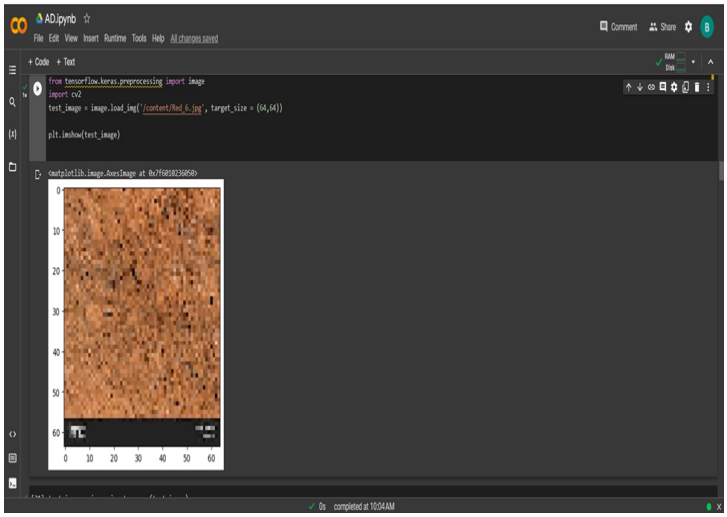
POOLING LAYER

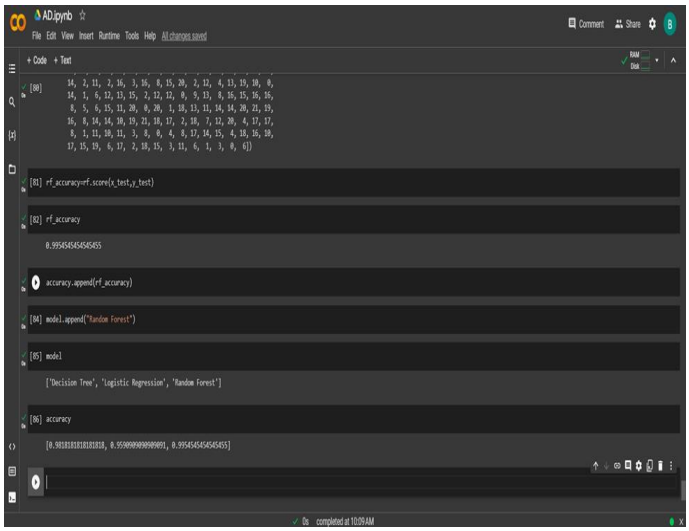
The (Max) Pooling Layer uses the 3x3 convolution layer input matrix and works to further reduce the dimensionality while retaining the key aspects of the image. We split the input into all

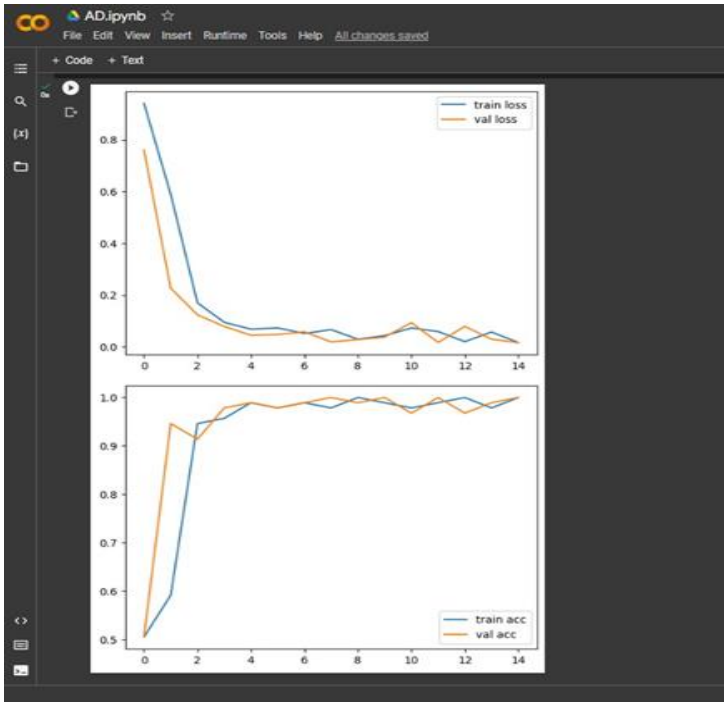


4. Results

The proposed model is based on soil and crop database. Several machine learning algorithms are used to classify the soil type. For a particular soil type suitable crop is suggested. From the experimental result ,We see that SVM has obtained the maximum accuracy. The classification accuracy is tabled below.







There is multiple distribution of crop suggestion under the condition of without irrigation and with irrigation. Rabi crops, Kharif crops 1, Kharif crops 2 are one kind of distribution of crops. Yearly crops and long lived crops are another type of distribution. Similar type of detailed crop suggestion of each class with specific map units are exists. Table VI of the following shows comparison result with some existing methods for soil classification. It shows that the proposed SVM based methods outperform than other methods mentioned here.

TABLE – Performance Comparision

Classification Model	**Accuracy (%)**	**Average Accuracy (%)**
Gaussian SVM	98.18	97.54
Weighted k-NN	95.90	
Bagged trees	99.54	

5. Conclusion

5.1 Project Conclusion

An application of classifying the soil types and providing the necessary crop suggestions for the classified soil series. The proposed work will benefit farmers to maximize productivity in agriculture, reduce soil degradation in cultivated fields, and reduce fertilizer use in crop production by recomending the right crop considering various attributes. The proposed work aids framers to accurately select the crop for cultivation and attain sustainability.

5.2 Future Scope

The purpose of soil classification system is to group together soil with similar properties or the attributes. The present model deals with available data whereas the future model contain the real time A data that is directly received from agricultural land that is placed with sensors. The sensors senses the soil fertility and other minerals contained in the soil.

6. References

V. Rajeshwari and K. Arunesh, Analyzing Soil Data using Data Mining Classification techniques, Vol 9(19),May 2016.

Jay Gholap , Anurag Ingole , Jayesh Gohil, Shailesh Gargade, Vahida Attar (2013), Soil data analysis using classification techniques and soil attribute prediction,.

Sk Al Zaminur Rahman, Kaushik Chandra Mitra ,S.M. Mohidul Islam(2018),Soil classification using Machine Learning Methods and Crop Suggestion based on Soil Series.

L.Armstrong , D.Diepevven & R. Maddern(2004),The Application of Data Mining Techniques to categorize agricultural soil profiles.

Chiranjeevi .M .N , Ranajana B Nadagoundar(2018), Analysis of Soil Nutrients using Data Mining Techniques.

Ramesh Vamanan, K.Kumar (2008),Classification of Agricultural Land Soils A Data Mining Approach.

Chandrakar PK , Kumar S, Mukherjee D(2011), Applying classification techniques in Data Mining in agricultural land soil.

Campus Valls G , Gomez Chova L , Calpe Maravilla J, Soria

Olivas E, Martin Guerreo JD, Moreno J(2003) Support vector machines for crop classification using hyperspectral data.

S

Bhuyar V(2014) , Comparative analysis of classification

techniques on soil data to predict fertility rate for Auranagbad District .

T .Mathavi Parvathi ,Automated soil testing process using combined mining processManonmaniam Sundaranar University.