

Implementation of Multilayer Neural Network with Decision Tree Model for Classification of Soil Type and suggesting suitable Crop Cultivation using Machine Learning Technique

A Zakiuddin Ahmed

Research Scholar, Department of Computer Science, Jamal Mohamed College (Autonomous)
Tiruchirappalli, Affiliated to Bharathidasan University, Tiruchirappalli, India

Dr. T. Abdul Razak

Department of Computer Science, Jamal Mohamed College (Autonomous) Tiruchirappalli,
Affiliated to Bharathidasan University, Tiruchirappalli, India

ABSTRACT

Soil is a significant element in agriculture. Several types of soil exist in different parts of our Country. Each type of soil has its own characteristic features and support growth of different kinds of crops. The yield of agriculture purely depends on environmental conditions and soil type. We need to know the features and characteristics of various soil types to understand which crop grows better in certain soil type. For the purpose of finding the classification of soil types the soil dataset is downloaded and based on the soil type predicted the farmers are suggested to cultivate the suitable crop. In this research work the various soil types are classified with the help of proposed algorithm using Multilayer Neural Networks with Decision Tree model. Experimental results illustrates the performance of generating best decision tree for classifying soil type from the given soil dataset. The algorithm for application of Multilayer neural networks with Decision Tree model helps to classify the soil types more accurately than the existing algorithms such as SVM, KNN, Bayesian approach to decision tree and Ensemble approach to Decision Tree model.

KEYWORDS: Decision Tree, Multilayer Neural network, Machine Learning, Soil Dataset

1. INTRODUCTION

Machine learning models play significant role in the agricultural data analysis. Machine learning model is one which learns from the historical data (Training data) fed to it and then build a mathematical model to predict the expected results for new set of data (Test data), which comes in as input to the system. A large quantity of dataset helps to build a better model, which predicts the desired output accurately.

Machine learning algorithm can be trained by supplying huge amount of data and then letting them explore the data, construct the logical model from the data and further predict the accurate results automatically without being explicitly programmed. The performance of the machine learning algorithms depends on the amount of data, which can be determined by cost function. To prepare the given datasets for the purpose of classification process, the data preprocessing techniques are applied.

Decision Tree is the supervised machine learning algorithm used for solving both classification and predication problems. Decision Trees are extensively used in agricultural sector to support decision making in selection of appropriate crop to be cultivated based on the soil types predicted. With the given datasets the Decision Trees are used to solve the business solution by using tree representation. In a Decision Tree, each internal node corresponds to an attribute and each leaf node corresponds to a class label. The soil datasets taken from a particular region are tested for the accuracy. The results of existing algorithms KKN, SVM, Decision Tree and Ensemble learning approach to decision tree model; and the proposed algorithm called multilayer neural network approach to the decision tree model are compared to show that the proposed algorithm produces more accurate results.

There are 8 major categories of soil that are found in India. They are alluvial soil, Black soil, Red soil, Laterite and Lateritic soil, Forest and Mountain soil, Arid and Desert soil, Saline and Alkaline soils, Peaty and Marshy soils. Maximum of 6 categories of soil might be present in a particular region.

In this research work, an algorithm for the proposed method has been developed and it is implemented using Python. The results obtained for the proposed algorithm and the results obtained from the existing algorithms are furnished in the Table 2 and visualization tools are used to compare and show the results visually.

2. RELATED WORKS

A.Zakiuddin Ahmed et.al [1] proposed a method using Ensemble learning approach to Decision Tree algorithm using machine learning technique for the purpose of classification of soil type using given set of large dataset of a particular region. Soil types are classified based on the chemical properties of the soil and finally suggesting the agriculturalists to cultivate the appropriate crops, which will give more profits to them.

Dr. D.Ashok Kumar et.al [2] performed a survey on data mining and pattern recognition techniques for classification of the soil type from the soil dataset. It was found that efficient methods need to be designed for the efficient classification of soil type to help the farmer to cultivate a suitable crop based on the type of soil, which will give them more financial benefit.

Vrushal Milan Dolas et al., [3] designed an improved decision tree algorithm to overcome the drawbacks found in the existing decision tree algorithms such as C4.5 and CART algorithms. The C4.5 is biased towards attributes with more values and CART algorithm produces misclassification errors when domain of the target attribute is very large. The proposed algorithm helps to accurately classify the Soil from the given large soil dataset of particular region.

Rushika Gadge et al [4] proposed the model that absorbed on verifying the soil quality to predict the appropriate crop to be cultivated according to the soil type. The proposed algorithm in this paper helps to classify the soil type excellently and based on the type of soil type, the farmer cultivate the suitable crop and maximize the yield based on the appropriate fertilizers recommended.

N.Saranya et al [5] experimented with machine learning classification algorithms such as K-Nearest Neighbor(KNN), Bagged tree, support vector machine (SVM) and logistic regression to predict the soil types from the given soil dataset. These algorithms were applied to mine the knowledge from soil data and two kinds of soil were considered such as Black and Red soil. Based on the soil type predicted, using the above said classification algorithms, the agriculturalists chose to cultivate the appropriate crop.

Chandan et.al [6] proposed a novel technique to classify soil type based on various soil features such as moisture, nutrients, structure, quality, pH and texture of soil using machine learning classification algorithms namely decision tree, Support Vector and K-Nearest Neighbor to classify the soil type from the given soil dataset.

Ashwini Rao et al.,[7] studied the significance of data mining tools and techniques and found out that there are several algorithms and techniques being used for classification of soil for the purpose of suitable crop cultivation. They studied existing Support Vector machine learning algorithm and enhanced the algorithm with the intention of the prediction of the accurate result while classifying the soil dataset. The proposed algorithms assimilate and process the colored image of soil sample, extract different features like color, texture etc. different soil type like red, black, clay, alluvial etc.

Sk Al Zaminur Rahman et al.,[8], designed an algorithm that predicts soil types with land type and according to prediction it can suggest the farmers about the suitable crops can be cultivated. Several machine learning algorithms such as weighted k-Nearest Neighbor, Bagged Trees, and Support Vector Machines based on Gaussian Kernel are used for soil type classification. Experimental results show that the proposed Support Vector Machine based method yields the best when compared with many existing methods.

J.Padarian et.al [9] have studied a novel method to predict soil properties from raw soil spectrum using the CNNs (convolution neural networks), a type of deep learning algorithm. The proposed method in this research paper proved to be effective when trained using large dataset. The high accuracy of CNN makes it an ideal tool for modeling soil spectral data.

3. PROPOSED MODEL

In the proposed model, an algorithm is designed using multilayer neural network with Decision Tree model which will more accurately classify the soil types than the already existing algorithms for the given soil dataset of a specific area. The proposed algorithm for classification of soil type using multilayer neural networks with Decision Tree model is designed and

it has been implemented by writing codes using python language- The results obtained from the proposed algorithm will be then compared with the results obtained for existing algorithms such as KNN, SVM, Decision Tree and ensemble learning approach to Decision Tree model.

A Multilayer Neural Networks with Decision Tree Classifier operates by breaking down a soil dataset into smaller subsets based on different criteria. The different sorting techniques will be applied to divide the dataset, with the number of examples getting smaller with each partition. It uses the divide and conquers method for splitting the data according to attribute values. Decision Tree analysis is utilized for predicting soil type. Decision Tree processing is a technique that includes the recursive partitioning of data into progressively homogeneous subsets. After the partitioning the subsets will be called as nodes. The label of the majority class is allocated to each terminal node.

Multilayer Neural Networks:

A Multilayer neural network is a deep learning technique used to solve the classification problem for nonlinear datasets by deploying the hidden layers, whose neurons are not directly connected to the output. Multilevel Perception (MLP) has multiple layers of neurons. The additional hidden layers can be interpreted geometrically as additional hyper-planes, which enhance the separation capacity of the network.

The network layers separation ability has been improvised by hidden layers addition and it can be geometrically interpreted. A deep neural network (DNN) is termed from artificial neural network (ANN) consists of several layers in-between input and output layers. For both linear and non-linear sets the expected output is arrived with the help of mathematical calculations. This deep learning framework corresponds to multilayer stack of simple elements performing linear and non-linear mappings. The inputs have been changed by the elements in the multilayer stack to selection increase and represent the invariance. For non-linear layers, complicated functions can be implemented and minute by minute detail can be found out.

The training occurs in a supervised style. The basic idea of this technique is to present the input vector to the network; calculate in the forward direction the output of each layer and the final output of the network. For the output layer the desired values are known and therefore the weights can be adjusted as for a single layer network; in the case of the BP algorithm according to the gradient decent rule. To calculate the weight changes in the hidden layer the error in the output layer is back-propagated to these layers according to the connecting weights. This process is repeated for each sample in the training set. One cycle through the training set is called an epoch. The number of epochs needed to train the network depends on various parameters, especially on the error calculated in the output layer.

Figure-1 shows typical multilayer network architectures with two and three layer networks,

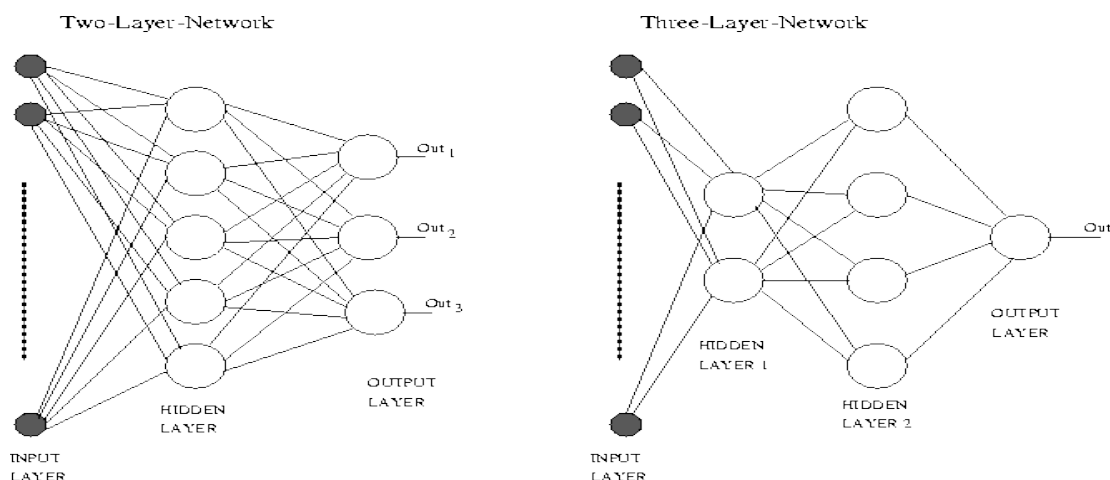


Figure -1: Example of Multilayer Neural Network

In this research paper, multilayer neural networks technique is used with decision tree model to improve the performance and accuracy of the result.

PROPOSED ALGORITHM

The proposed algorithm shows the classification of soil type using Multilayer Neural Network with decision tree model in the machine learning mechanism. For this research paper soil dataset with the name “soil.csv” which consists of 5200 records and 13 attributes, is downloaded and it has been used as input to the existing algorithms and as well as in the proposed algorithm. Initially, 80%, i.e., 4160 records of the training sample or training dataset is given and trained to build a classifier model for the prediction and 20%, i.e., 1040 records (unknown class label) of the testing data are used for the process of soil type prediction.

The following program code is written using python language to split the soil dataset into training sample and testing data:

```
import pandas as pads;
from sklearn. model_selection import train_test_split;
new_data = pads. read_csv(“soildata.csv”);
new_data. Head ();
x_train, x_test, y_train, y_test = train_test_split (x,y, test_size=0.2);
x_train. head ();
```

The classification process has been done with the help of the combination of multilayer neural networks with the decision tree model in the machine learning algorithm. The classes are defined to predict the attribute value for the soil prediction. The steps for soil classification using multilayer neural networks with decision tree model is given below

Step 1: The input is given as dataset ‘DS’ which is considered as sample soil dataset for soil type prediction

Step 2: The given soil datasets are preprocessed for our research need.

Step 3: The feature extraction process is performed, which will reduce the dimensionality of the data given as input. Then datasets are trained and tested.

Step 4: Now the Multilayer neural network techniques are used with decision tree for the purpose of the classification of the given soil dataset.

Step 5: Finally, tree formation is done and soil type is predicted by using the performance analysis carried out using the parameters precision, recall, F1-Score and accuracy.

4. RESULTS AND DISCUSSION

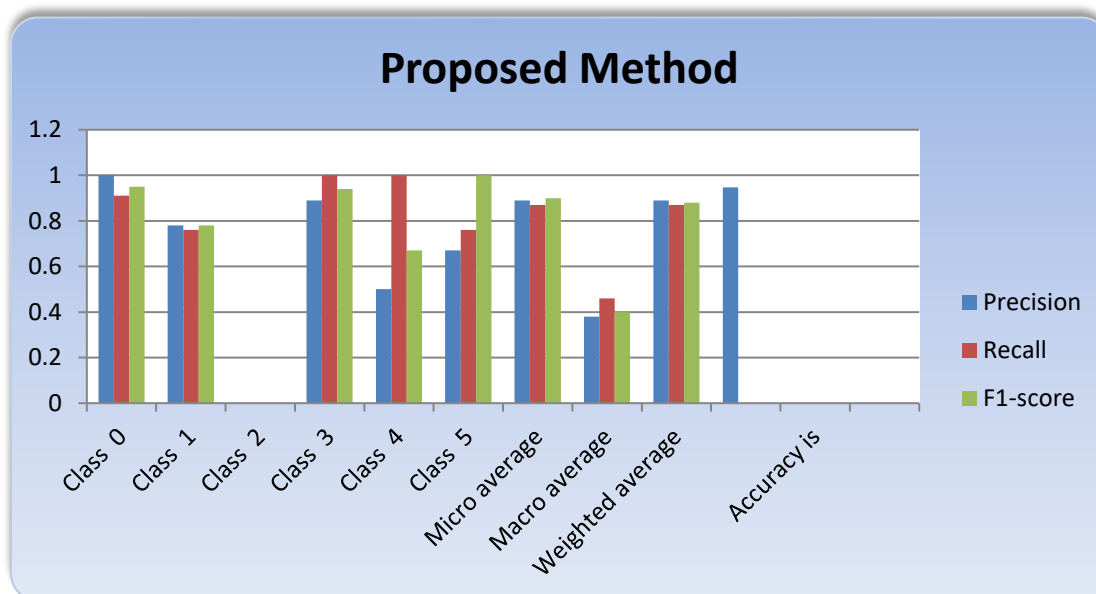
The proposed algorithm for classification of Soil type using multilayer neural network with Decision Tree Model is implemented using the Python and the results are furnished in the Table- 1. The proposed algorithm uses the measures such as Precision, Recall, F1-Score and accuracy. These metrics are connected with mining classification of soil type from the soil dataset of a specific region. These metrics are the base for making the judgment of the classification task.

Table 1: Performance analysis of proposed technique

Soil Type	Precision	Recall	F1-score	Support
Class 0 (Red Soil)	1.00	0.91	0.95	395
Class 1 (Black Soil)	0.98	0.86	0.88	127
Class 2 (Sandy Soil)	0.00	0.00	0.00	0
Class 3 (Literate Soil)	0.89	1.00	0.94	338
Class 4 (Red loamy Soil)	0.56	1.00	0.69	93
Class 5 (Alluvial Soil)	0.77	0.76	1.00	87
Micro average	0.89	0.87	0.90	1040
Macro average	0.38	0.46	0.40	1040
Weighted average	0.89	0.87	0.88	1040

Accuracy	0.957368	
-----------------	-----------------	--

Figure-2: Graphical representation of proposed Method



The graphical representation of the results obtained using the proposed algorithm is shown in the Figure-2. There are six classes which are classified in the proposed algorithm namely class 0(Red Soil), class 1(Block or regular soil), class 2(Sandy Soil), class 3(Red sandy soil), class 4 (Gray Soil) and class 5(Gray Sandy Soil). The overall accuracy is 95% which is predicted from the proposed algorithm.

Based on the soil types classified, suggestions are given to the farmers about the suitable crop that they can cultivate. The below Table-2 shows the crops against each soil type.

Table- 2: Suitable Crops against the soil types

Soil Type	Suitable Crop
Class 0	Cotton, Rice, wheat, Bajra, Groundnut, black gram, green gram, maize, all oil seeds, Vegetables and fruits.
Class 1	Rice, Sugarcane, Sunflower, cereal crops, citrus fruits, tomatoes, ground nets and all oil seeds.
Class 2	Rice, Wheat, Ragi, groundnut, potato, pulses, oil seeds and fruits such as mango, orange, citrus etc.
Class 3	Cotton, rice, wheat, pulses, tea, coffee and rubber
Class 4	Wheat, cotton, corn (Maize), Millets, pulses and barley
Class 5	Tea, Coffee, Spices, Wheat, Barley, tropical fruits.

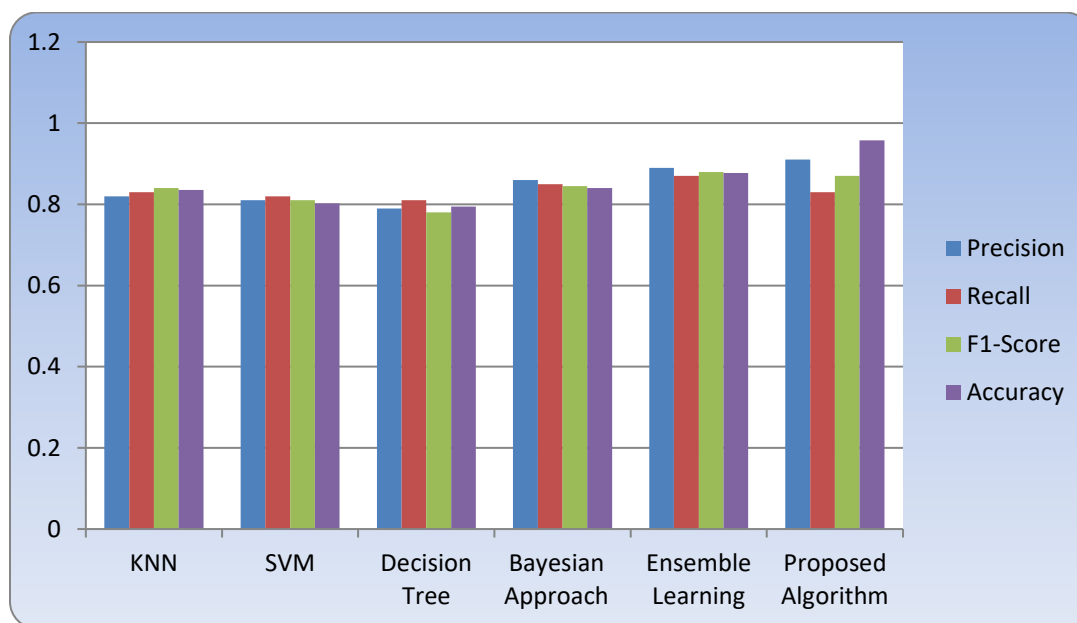
Comparative Analysis of Existing Techniques with Proposed Model

The Comparative analysis has been performed here by comparing the results obtained in the existing methods with the proposed algorithm and the values are furnished in the below table 3 and the same is visually shown using bar chart in the following figure -2.

Table- 3: Comparison of the existing techniques with proposed model

Algorithms	Precision	Recall	F1-Score / F1 measure	Accuracy
K Nearest Neighbor Algorithm	0.82	0.83	0.84	0.8356872
Support Vector Machine Algorithm	0.81	0.82	0.81	0.802529
Decision Tree	0.79	0.81	0.78	0.794765
Bayesian Approach	0.86	0.85	0.84521	0.840508
Ensemble Learning Approach	0.89	0.87	0.88	0.876953
Proposed Algorithm	0.91	0.83	0.87	0.957368

Figure-3: comparison of the existing algorithms with proposed algorithm



The comparison of the existing algorithms such as K-Nearest Neighbor (KNN), Support Vector Machine (SVM) , Decision Tree (DT), Ensemble Learning Approach to Decision Tree classification model with the proposed algorithm named classification of soil type using multilayer neural networks with Decision Tree model is shown in the Table 3. The same is visually depicted with help of Bar-chart in the Figure 3. Here the proposed algorithm has a better precision, recall and F1-score and accuracy than the existing algorithms taken for this research paper.

5. CONCLUSION:

Classification of soil type has been performed successfully and accurately for the given datasets of a particular region, which would be helpful for the cultivation of suitable crop by the farmers. The professional methods are adopted to perform the soil classification from the complex soil datasets. In this paper, the use of multilayer neural network with decision tree model is adopted to predict the soil type using machine learning techniques. Here, analysis of soil type using different classification algorithms such as K-Nearest Neighbor, Support Vector Machine, Bayesian approach to Decision

Tree, Ensemble learning approach to Decision Tree Algorithm and the proposed method use of multilayer neural networks with Decision tree model have been compared. Finally, the proposed algorithm gives the better result than the existing algorithms said above.

References:

1. A.Zakiuddin Ahmed and Dr.T.Abdul Razak, "**Ensemble Learning Approach to Decision Tree Algorithm for the Classification of Soil type and Suggesting Suitable Crop Cultivation Using Machine Learning Techniques**" Shodha Prabha (UGC CARE Journal), Vol.47, Issue.01, No.3:2022, ISSN:0974-8946
2. Dr. D. Ashok Kumar, N. Kannathasan, "**A Survey on Data Mining and Pattern Recognition Techniques for Soil Data Mining**", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 1, pp. 422-426, May 2011, ISSN (Online): 1694-0814
3. Vrushal Milan Dolas, Prof.Uday Joshi, "**A Novel Approach for Classification of Soil and Crop Prediction**", International Journal of Computer Science and Mobile Computing (IJCSMC), Vol.7, Issue 3, pp.20-24, March 2018, ISSN 2320-0887.
4. RushikaGadge, JuileeKulkarni, Pooja More, Sachee Nene, Priya. R, "**Prediction of Crop yield using Machine Learning**", International Journal Research Journal of Engineering and Technology (IRJET), volume:05, issue: 02, pp. 2237-2239, Feb 2018, e-ISSN: 2395-0056.6.
5. N.Saranya and A.Mythili, "**Classification of Soil and Crop Suggestion using Machine Learning Technique**", International Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 02, pp. 671-673 February-2020, ISSN: 2278-0181
6. Chandan I, Ritula Thakur, "**Recent Trends Of Machine Learning In Soil Classification: A Review**", International Journal of Computational Engineering Research (IJCER), Volume, 08, Issue, 9, pp. 25-31, September – 2018, ISSN (e): 2250 – 3005.
6. Ashwini Rao, Janhavi V, Abhishek Gowda N.S and Mrs.Rafega Beham, "**Machine Learning in Soil Classification and crop detection**", International Journal of Scientific Research & Development, Volume:4, Issue: 1, pp.792-794, April 2016, ISSN (online): 2321-0613
7. Sk Al Zaminur Rahman, Kaushik Chandra Mitra, and S.M. Mohidul Islam, "**Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series**", 21st International Conference of Computer and Information Technology (ICCIT), Date of Conference: 21-23 Dec. 2018, Date Added to IEEE Xplore: 04 February 2019, ISBN Information: INSPEC Accession Number: 18432260, Publisher: IEEE
8. J.Padrian, B.Minasny and A.B.McBratney, "Using Deep Learning to predict soil properties from regional spectral data", Geoderma Regional Publication, Volume 16, March 2019, e00198