Soil Classification and Crop Prediction Using Machine Learning

Yuvraj Jangir(✉), Tushar Goyal, Sumit Kandari, and Arshad Husain

Department of Computer Science, DIT University, Dehradun, Uttarakhand, India
Yuv.rraj786@gmail.com

Abstract. Soil classification is the process in which soil is segregated according to its physical and chemical properties. This process can be achieved manually or using a machine learning algorithm. The use of machine learning algorithms has been on the rise in recent years due to their accuracy. They can classify soils with more precision than humans can manually, by considering many factors such as pH, organic matter content, and particle size distribution. We here proposed a model to classify soil and to predict the most suitable crops using various algorithms of machine learning like Convolutional Neural networks (CNN), Decision Trees, Naive Bayes. Soil and crop datasets are used, they comprise of different geographical and physical. Parameters. The module is tested on manually created datasets and results are obtained.

Keywords: Soil types · Machine learning · Convolutional neural network (CNN) · Decision tree classifier

1 Introduction

1.1 Objective of Study

To alleviate the agricultural crisis in its current state, it is necessary to put in place better recommendation systems to alleviate the crisis by helping farmers make informed decisions before planting begins. The main objective of this experiment was to classify and predict the suitability of different types of soils on various crops based on predictor variables like temperature, rainfall, and location. Prediction obtained using decision trees algorithm. It is important to note that this algorithm works only when a set number of input values are set up in the form of training samples, which gives a result prediction for each sample in accordance with that input setting, which can then be compared against ground truth at other input settings (some samples may not have the rule applicable). As such, each predictor variable was considered as a feature to define the suitability of a given map. A decision tree was then constructed to mimic this process and so predict the result for an input location.

R. Mehra et al. (Eds.): ICCISC 2022, CCIS 1672, pp. 16–21, 2022.
https://doi.org/10.1007/978-3-031-22915-2_2

1.2 Related Works

In order to predicts soil type and based on the prediction, suggests suitable a model is proposed which includes several machine learning algorithms are used for soil classification. Experimental results show that the proposed SVM (support vector machines) based method performs better than many existing methods [1]. Soil dataset and crop dataset were used to classify the soil. Soil dataset contains class labelled chemical feature of soil. The crop suggestion dataset contains class labelled crop suggestion attributes [2]. The soil image has been analyzed using various image preprocessing. Soil color is identified using statistical properties such as mean of Red, Blue, Green (RGB) values of image pixels. A classification of soil based on feature extraction of soil color, soil pH values and texture by using Support Vector Machine classifier. Here, the one-against-all SVM method is used for classification. Recommended which crops were suitable for tested soil image based on feature extraction by using image processing [3]. Digital image processing and Image analysis technology has been used in which suggestions of crops to be grown in that soil type. The image of the soil/land is clicked using phone camera and submit it. An Image is a two - dimensional signal. Image processing is a method to perform some operations on an image, to get an enhanced image or to extract some useful information [4]. A model has been proposed for predicting the soil type and suggesting a suitable crop that can be cultivated in that soil. The model predicts soil fertility and 8 other properties of agricultural land [5]. A comprehensive system is described for soil classification in which different images of soil samples are captured. The features of each type of soil are collected and are stored in a separate database. This database is later used in the final stage for soil classification [6]. In order to efficiently classify the soil instances and maps the soil type to the crop data to get better prediction with higher accuracies. Soil prediction involves types of crop classifications and geographical attributes. It also aims at creating a system that processes the real-time soil data to predict the crops with higher accuracy [7].

2 Proposed Framework

In present work, we propose a novel methodology for soil classification using Convolutional Neural Network (CNN). (Fig. 1.) The generated classifiers were validated with the accuracy of more than 90%. Secondly, we discussed about crop prediction using decision tree-based classifier.

In this paper, we described a proposed architecture for soil classification using image samples. Soil classification is done by analyzing different types of soil image samples. Enough training samples are needed to classify soil samples. Training samples should be selected carefully. After selecting the training dataset, all the images are then passed through Image Processing which includes resizing and augmentation process. After applying Image Processing on all images, the resultant images are passed through Convolutional Neural Network (CNN) for classification.

Convolutional Neural Network (CNN) Is a powerful algorithm for image processing. These algorithms are currently the best algorithms we have for the automated processing of images. CNN is a directed acyclic graph with four main layers, which are: input layer,

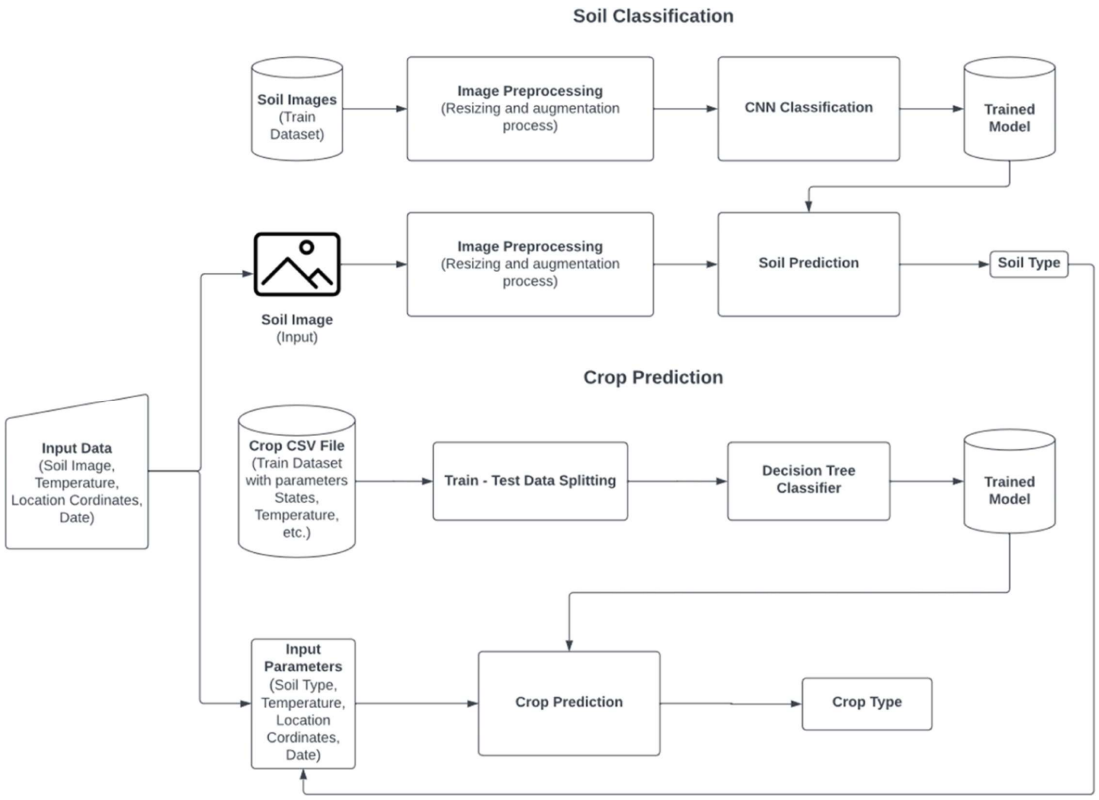

Fig. 1. Architecture of the system

filter layer (convolutional layer), pooling layer and output layer to provide better accuracy of classification. CNN is used to classify the soil samples by extracting various features from the image. (Fig. 2.) The dataset used in the paper contains nearly 200 cropped images from different soil types. It is a type of machine learning algorithm that allows for classification and prediction tasks. It is used to classify inputs into several categories and predict an output based on given inputs. Images contain data of RGB combination. The computer does not see an image, all it sees is an array of numbers. Color images are stored in 3-dimensional arrays. The first two dimensions correspond to the height and width of the image (the number of pixels). The last dimension corresponds to the red, green, and blue colors present in each pixel.

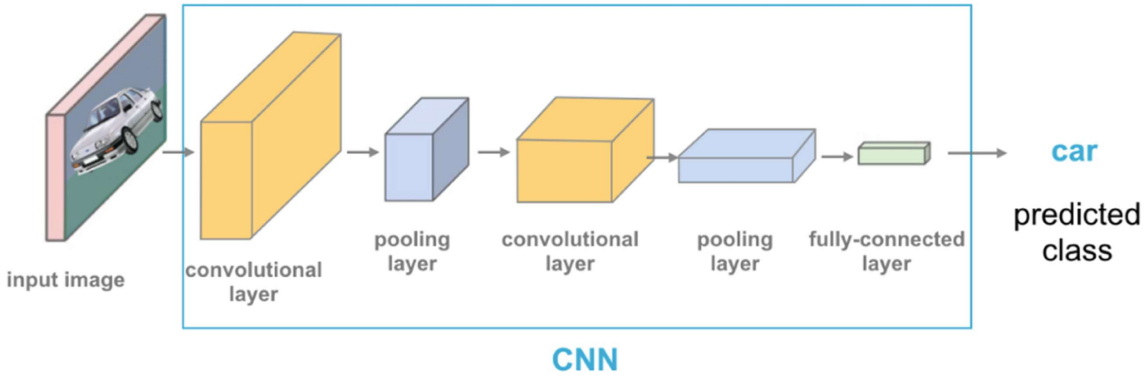

Fig. 2. Working of CNN

The output of CNN is compared with a threshold value and if it is greater than the threshold value then it will be classified into class1 otherwise it will be classified into class2. The classification process is done using Python 3.10.4 and Keras 2.8.0 lucid framework. The evaluated accuracy of this method is greater than 90%, which means the method is reliable, accurate and fast. Then, the trained model is stored in h5 format, which is then used to predict the soil, using the image provided in the input.

In the crop prediction model, we first created a dataset with the following parameters: states, rainfall, ground water, temperature, soil type, season, and crop, and stored it in a csv file. After creating the dataset, we split the data for the training data set and the testing data set separately. After that, we used the decision tree classifier to train the crop prediction model. The performance of the model was measured by the accuracy value.

Decision Tree Is a supervised prediction method which is widely used because they can easily obtain input data and are suitable for classifying the data into a wide range of categories. Decision trees can be used as a type of classifier or regression model that uses binary trees to predict outcomes. The key advantage of decision trees is that they can be easily implemented and interpreted. Decision tree models have been shown to be effective in many real-world applications. Decision tree classifiers are used in many fields, such as pattern recognition, data mining, machine learning and bioinformatics. Decision trees can be used to predict categorical outcomes, and they frequently have an option to do regression functions.

In this study, we presented a real-world dataset that has the names of various crops. We also had the opportunity to determine two meteorological inputs: rainfall and ground water level. These two inputs were necessary to help us build a crop prediction model using decision trees. After training the model, we stored the model in a.sav format, which is then used to predict the crop, using the parameters provided in the input and the soil type predicted using soil classification model.

3 Result and Discussion

3.1 Dataset Collection

Multiple datasets are used to train and obtain relevant results. All the datasets used are custom datasets; built and structured according to the requirements of the algorithm and the proposed test cases. Below is the list and types of the datasets used-

1. *Soils.zip (Soil Image Dataset)* - Contains about 150–200 images of different types of soil which are used for agriculture and found in the Indian subcontinent.
2. *Cat_crops.csv* - The CSV file mentioned contains data on various parameters that were considered when training the machine learning model for the crop recommendation system.

```
19/19 [==============================] - 5s 281ms/step - loss: 0.4118 - acc: 0.8564
Epoch 19/30
19/19 [==============================] - 5s 269ms/step - loss: 0.5900 - acc: 0.7790
Epoch 20/30
19/19 [==============================] - 5s 259ms/step - loss: 0.4046 - acc: 0.8232
Epoch 21/30
19/19 [==============================] - 5s 276ms/step - loss: 0.4455 - acc: 0.8453
Epoch 22/30
19/19 [==============================] - 5s 257ms/step - loss: 0.2921 - acc: 0.8508
Epoch 23/30
19/19 [==============================] - 5s 265ms/step - loss: 0.3423 - acc: 0.8508
Epoch 24/30
19/19 [==============================] - 5s 260ms/step - loss: 0.4867 - acc: 0.8508
Epoch 25/30
19/19 [==============================] - 6s 294ms/step - loss: 0.2686 - acc: 0.9061
Epoch 26/30
19/19 [==============================] - 5s 260ms/step - loss: 0.1788 - acc: 0.9227
Epoch 27/30
19/19 [==============================] - 5s 263ms/step - loss: 0.4485 - acc: 0.8343
Epoch 28/30
19/19 [==============================] - 5s 268ms/step - loss: 0.1840 - acc: 0.9392
Epoch 29/30
19/19 [==============================] - 5s 267ms/step - loss: 0.3575 - acc: 0.8674
Epoch 30/30
19/19 [==============================] - 5s 262ms/step - loss: 0.2099 - acc: 0.9227
```

Fig. 3. Result obtained

3.2 Results

The below Fig. 3. Represents the accuracy of the soil classification model which was built using CNN (Convolutional Neural Networks).

After several changes and observations, it has been noticed that after improving the dataset, the accuracy of the model is also improved.

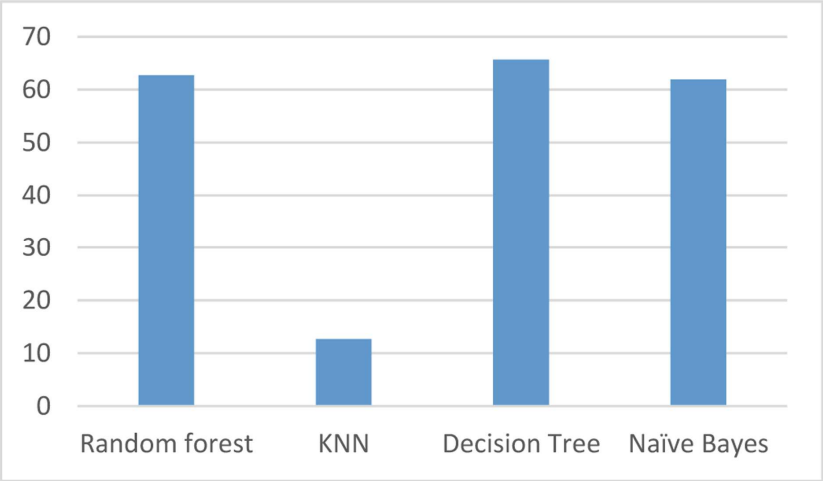

Fig. 4. Accuracy of different algorithm

In crop prediction model, we provided a csv file which consists of the following parameters: States, Rainfall, Ground Water, Temperature, Soil Type, Season and Crop. We compared the results with the previous prediction methods. (see Fig. 4.) We used 4 different classifiers to train the crop prediction model, out of which the decision tree

classifier gave us the best accuracy among them. The accuracies of all the different classifiers used were:

4 Conclusions and Future Scope

This proposed system is based on an image processing technique where digital images of the soil samples were processed using convolutional neural network (CNN). In this study, Decision Tree was used to determine the crop suitability of the soil sample. The results showed that CNN recommended which crops were suitable for tested image of the soil samples. So, the proposed method will help farmers to increase the productivity of yield by identifying suitable crops for the soil samples.

In future perspective, a point location-based rainfall prediction and ground water detection module can be integrated with the other parameters. This would increase the overall prediction of suitable crops. Also, the dataset for all these three experiments consisted of nearly 200 cropped images from different soil types, a larger dataset is needed. This will increase the accuracy of prediction and classification of soil as well as crop.

References

1. Rahman, S.A.Z., Mitra, K.C., Islam, S.M.M.: Soil classification using machine learning methods and crop suggestion based on soil series. In: International Conference of Science and Technology Computer (2018)
2. Reddy, K.M.A., Chithra, S., Hemashree, H.M., Kurian, T.: Soil classification and crop Yadav, suggestion. Int. J. Res. Appl. Sci. Eng. Technol. (2020)
3. Yadav, P., Ahire, P.: Soil health analysis for crop suggestions using machine learning. AEGAEUM J. (2020)
4. Aishwarya, M., Revathy, R., Periasamy, J.K., Srujana, T.: Soil classification and crop suggestion using machine learning techniques. J. Gujrat Res. Soc. (2019)
5. Saranya, N., Mythili, A.: Classification of soil and crop suggestion using machine learning techniques. Int. J. Eng. Res. Technol. **9**, 671–673 (2020)
6. Chandan, R.T.: An intelligent model for indian soil classification using various machine learning techniques. Int. J. Comput. Eng. Res. (IJCER) **33**, 3005 (2018)
7. Shravani, V., Uday Kiran, S., Yashaswini, J.S., Priyanka, D.: Soil classification and crop suggestion using machine learning. Int. Res. J. Eng. Technol. (IRJET) (2020)