e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

SOIL CLASSIFICATION AND CROP SUGGESTION USING ML

Vidya Tulaskar*1, Rita Dhoble*2, Divya Bawane*3, Sanjana Virmuttu*4, Komal Ambagade*5, Abhimanyu Dhutonde*6

1,2,3,4,5 Students Department of Computer Science & Engineering, Tulsiramji Gaikwad-Patil College Engineering, Nagpur, Maharashtra, India

*6 Professor (Guide), Department of Computer Science & Engineering, Tulsiramji Gaikwad-Patil College Engineering, Nagpur, Maharashtra, India

ABSTRACT

India is an agricultural nation and one of the top three global producers of numerous crops. Despite being in the center of the agricultural industry, the majority of Indian farmers continue to be at the bottom of the social strata. In addition, despite the limited technical options available today, farmers still struggle to select the crop that is both financially and agronomically profitable for their soil because soil types vary widely around the globe. In order to predict the best crop to be cultivated, this research provides a crop recommendation system that makes use of a Convolutional Neural Network (CNN) and a Desultory Forest Model. These parameters include the region, soil type, yield, selling price, etc.

Keywords: Soil, Crop, Agriculture, crop recommendation, soil classification, machine learning.

I. INTRODUCTION

Two-thirds of the Indian population directly depend on agriculture for their living, consequently it has traditionally been and perpetuates to be one of the key substrata of the Indian economy. Adscititious consequential is the fact that it accounts for 20% of India's GDP (GDP). The farmer, who accommodates as our nation's Anna data (Victuals Provider), is at the center of the agricultural industry and is now dealing with a number of challenges: Use the enter key to start a new paragraph. The appropriate spacing and indent are automatically applied.

1) Farmers traditionally find it difficult to choose which crop is most suited and financially advantageous to their soil, their circumstances, and their location because of the variability in soil types across the country, and as a result they frequently suffer losses.

2) Due to unpredictable weather patterns, it is currently astronomically difficult for farmers to predict the yield for a specific planting season and the profit that they may make.

3) Due to the "farm to market" system, which involves hundreds of intermediaries who victual up the majority of the revenues by transporting and selling crops, farmers get dismally little returns for their labor.

Artificial intelligence and machine learning are widely used in modern agriculture. Precision farming methods, crop recommender systems, and yield prediction methods may all be used to improve farm output, plant pest detection, and overall harvest quality. AI system deployment might provide the struggling agriculture industry a boost. One of the emerging technologies in agriculture is machine learning. The agriculture industry may use machine learning to improve crop quality and output. Finding trends in the agricultural data and relegating them to more important data might become habitual. You can use this information for further procedures. Gathering data, processing it, and training it before testing it on samples of data are the steps that machine learning approaches often take. For the relegation of soil and crop prognostication based on prior patterns followed and the kind of soil, an algorithm like SVM may be used. The following datasets are needed for the project: a soil dataset with multiple chemical qualities and a crop dataset with geographical information.

Greema S Raj et al. [1] in this paper, the author feels that thescientists who are working on different land areas and relegating different soil types should have a prevalent language to ken about that particular soil. They opted for machine learning and some of the well-kenned algorithms fortheir research work. There are a variety of soil present in theworld having different characteristics and features and different crops can be grown on them. For the relegation of the soil Decision Tree, Neural networks, Naïve Bayes, and SVM techniques are utilized for comparative study purposes.

Mohammad Diqi et al. [2] this work is all about relegating thesoil type ebony and red, and for that purport, authors have utilized CNN and deep learning algorithms to relegate the resulting output. In this process, they got a prosperity rate of 97% and a loss of 0.1606.

Pallavi Srivastava et al. [3] in two streams, this study discussesseveral computer-based soil categorization techniques. The first kind of methodology uses image processing and computer vision to categorise soil using various parameters including texture, colour, and particle size, as well as more traditional image processing algorithms and techniques. Thesecond category includes soil categorization methods based on deep learning and machine learning, such CNN, which produce cutting-edge outcomes. By streamlining the entire process, deep learning applications primarily reduce the reliance on spatial-form designs and preprocessing methods.

Shravani V et al. [4] the goal of the research is to develop a model that effectively categories soil instances and maps the soil type to crop data to produce better predictions with higher accuracy. Crop classifications and regional characteristics areboth used in soil prediction. In order to anticipate crops moreaccurately, it also seeks to develop a system that analyses real-time soil data. There are two phases to the model: the training phase and the testing phase. Soil and crop databasesare the two utilized datasets. The list of the proper classes is generated after comparing the expected and actual classes.

S. A. Z. Rahman et al. [5] in this proposed work, authors provide a model that predicts soil series with regard to land type and, in accordance with prediction, suggests appropriatecrops. For soil classification, a number of machine learning techniques are utilised, including weighted k-Nearest Neighbor (k-NN), bagged trees, and Support Vector Machines (SVM) based on a Gaussian kernel. The suggested SVM-based technique outperforms several current methods,according to experimental data.

II. METHODOLOGY

1)k-nearest neighbors:

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

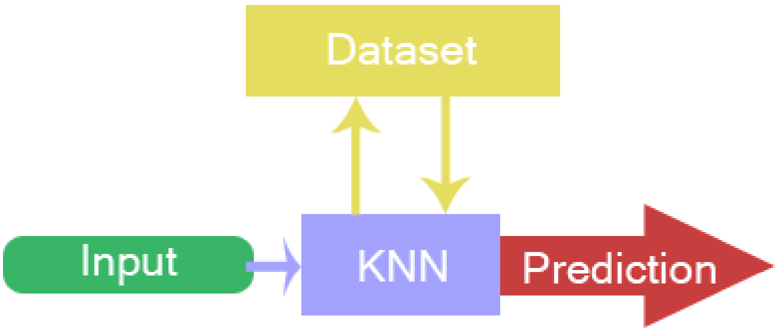


Figure 2.1: System Architecture of KNN

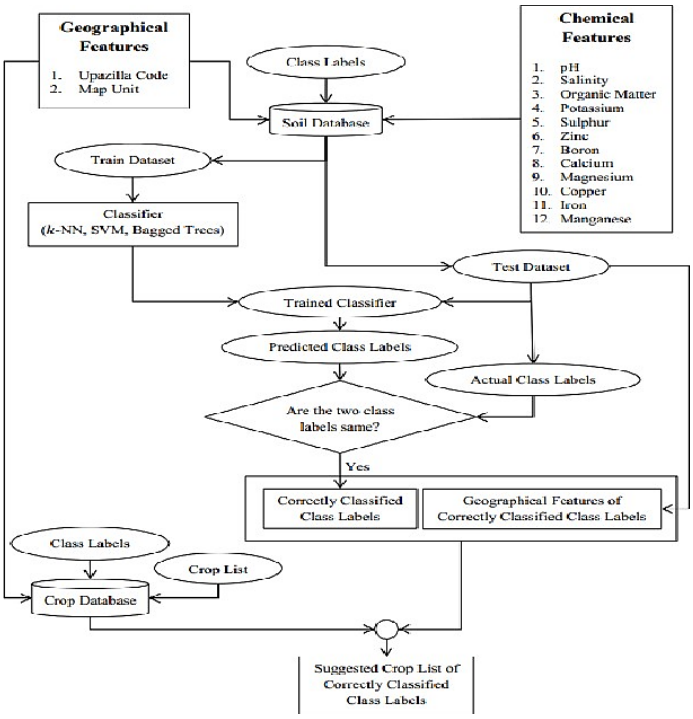


Figure 2.2: Proposed System Flow

i. Dataset Collection**:**

Images of the many soil types, including Red Soil, Black Soil, Clay Soil, and Alluvial Soil, are included in the dataset used to train the parameters of thismethod. Each kind of soil is represented by 150–200 photographs in the training set and 50–60 images in the test set of the dataset. This information is gathered from relevantinternet sources, including the Soil Classification Image Dataset on Kaggle.

ii. Data Pre-Processing and Algorithm Implementation:

Since the photos in the dataset are different shapes and sizes,they must first be pre-processed and scaled before being fed into the model. For our use case, the algorithm must read a colored picture, and every colored image in RGB format hasthree channels—one for each of the three colors—Red, Green, and Blue. Thus, 300x300x3 pictures are initially created from the original photographs. Then, in order for themodel to handle the data, the pictures are transformed into numerical (pixel intensities) arrays. The machine learning model is then given this numerical data.

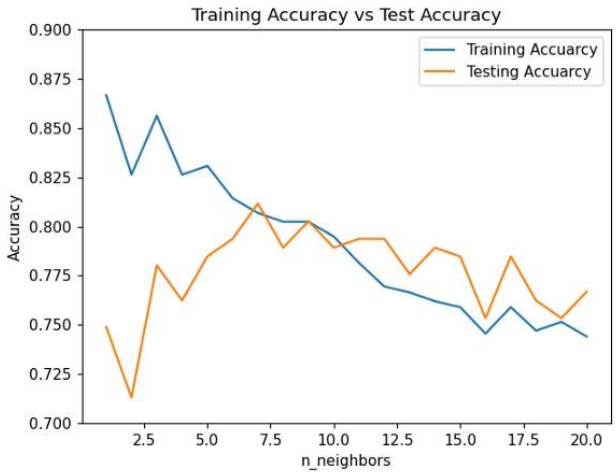


Figure 2.3: Graph of Accuracy of KNN vs No. of Training

With the learning rate set for 20 epochs and the AdamOptimization Algorithm, this network was trained. On the test set, the model had an accuracy of 95.21 percent. The Accuracy and Loss graphs in Figure show the training outcomes for this model.

2) Support Vector Machine (SVM):

The photos are initially downsized to 200 × 200 x 3 images using this method. The pictures are next transformed into numerical (pixel intensities) arrays, and finally flattened to afeature vector of 120000 by 1. The SVM model is then giventhis feature vector. Two kernels, the linear and the Radial Basis Function (RBF)Kernel, were used to train the SVM model. On the test set, the accuracy of the models with linear and RBF kernels was 83.5% and 86.7%, respectively.

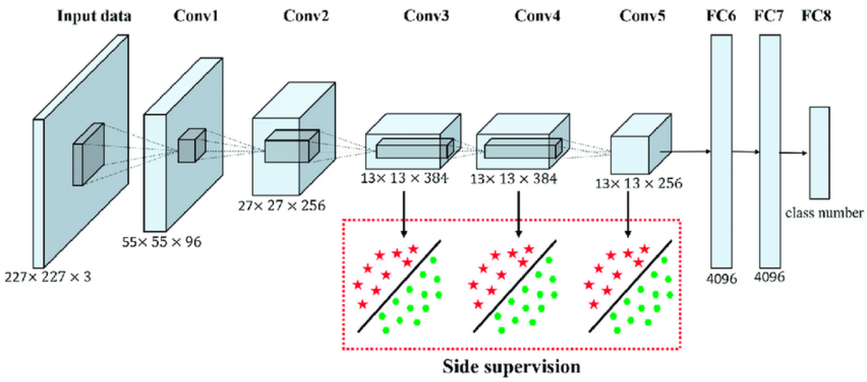


Figure 2.4: SVM Architecture

AlexNet: The Convolutional Neural Network Algorithm usesthis traditional network design. The photographs were initially downsized to 227 × 227 x 3 images in order toapproximate the proportions for this design. The CNN

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

modelis then given the numerical data from the arrays of pixelintensities that were created from the transformed pictures. Since we only have 4 classes, we only have 4 units rather than1000 units. As a result, there are around 59 million trainableparameters in the network. The Adam OptimizationAlgorithm was used to train the model, using a learning rateof 10-3 and an epoch size of 25. On the test set, thisarchitecture provided us an accuracy of 25%.

3)Convolutional Neural Networks with Other/New Architectures:

We chose to test several CNN architectures after receiving satisfactory results from the SVM model and dismal ones from SVM. The photos are initially downsized to 300x300x3for all the ensuing architectures. The CNN model is then supplied with the numerical (pixel intensities) arrays createdfrom the pictures. The Adam optimization technique is used to train the model, with a learning rate of all architectures anda total of 25 training epochs. A brief step by step procedure of designing the crop recommendation system is explained as follows:

Step 1: The soil dataset, which must go through pre- processing, is included in a comma separated values file.

Step 2: Pre-processing methods used on the input dataset include filling in missing values, encoding categorical data, and scaling values to the right range.

Step 3: Using the given split ratio, the pre-processed datasetis then divided into a training and testing dataset. The suggested work uses a split ratio of 75:25, meaning that 75% of the dataset is utilized to train the ensemble model and the remaining 25% is used as a test dataset.

Step 4: Individual classifiers are built using the training dataset. Each independent base learner receives the training dataset, and each classifier is created using each of them.

Step 5: Analyzing the data with each classifier the individualclass labels are acquired when each classifier applies the testing dataset.

Step 6: Assembling the individual classifier output using Majority Voting Technique the final result of soil classification and crop suggestion can be achieved.

III. MODELING AND ANALYSIS

Python: Python is a trending technology these days and provides multiple ways to use it on daily basis. It can be integrated with many languages like machine learning, data science, data analytics, Internet of Things, Artificial Intelligence, etc. It is an interpreted language and hence easyfor the beginners to understand and work on. In this project python is used as a core language.

Pycharm IDE: It is an Integrated Development Environmentcommonly known as IDE work on python programs and codes. This software allows user to implement their ideas andget the desired output by integrating anaconda navigator environment in it.

Anaconda Navigator: It is a desktop GUI (Graphical User Interface) application to work with jupyter notebook,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

launchapplications, providing conda environments, without using CLI (Command Line Interface). It has feature of collecting different python libraries.

Gaussian Kernel based SVM: SVM divides class objects into several decision planes. The difference between an itembelonging to one class and another is determined by a decision boundary. The data points that are closest to the hyper-plane are known as support vectors. The inputs are transformed into a higher dimensional space by the kernel function, which then separates the nonlinear data. They employed the Gaussian kernel function in their research. SVM was used, and the accuracy was 95.99%.

IV. RESULTS AND DISCUSSION

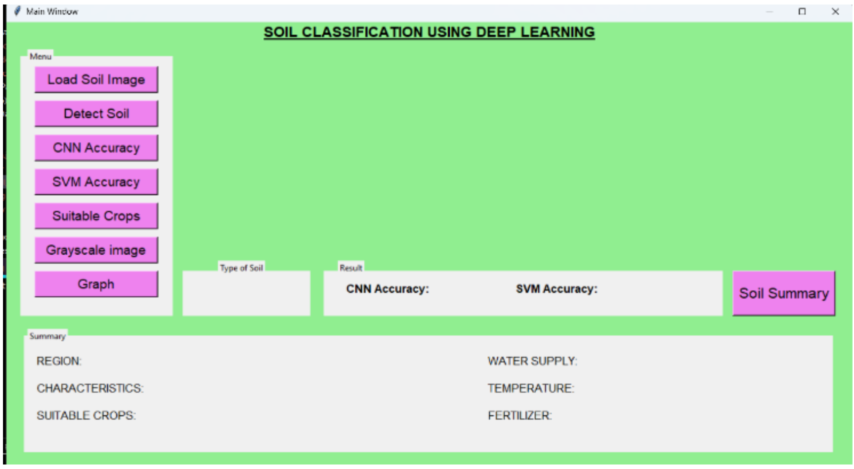
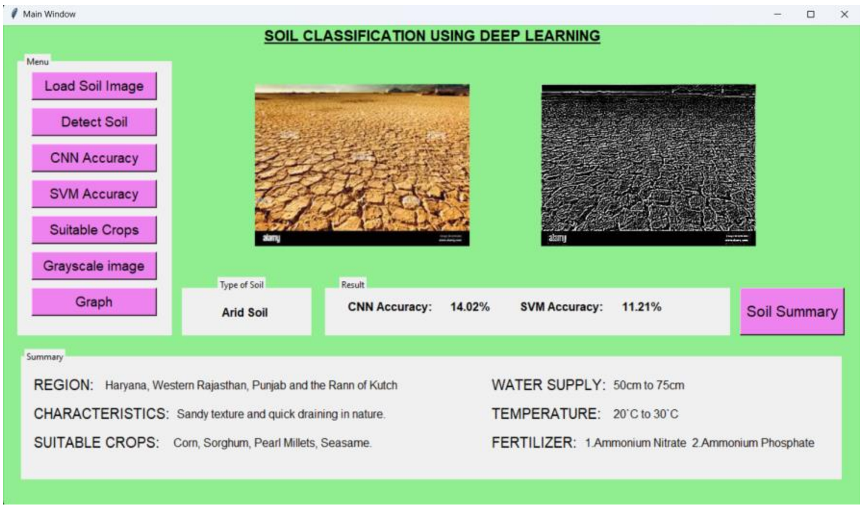


Figure 4.1: GUI of the proposed system

In the above given Figure 4.1 image, a GUI based on Python tkinter library is shown where we can see a dashboard with multiple buttons and some group of labels expected to show their final working with the correspondence of background task implemented. Each button in this GUI is having the particular task assigned for e.g. first button has the working of uploading the image from the browser by the user and to load it in the GUI for further processing. All the buttons likely to accessed according to the user preferences.



e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:06/June-2023 Impact Factor- 7.868 www.irjmets.com

Figure 4.2: Overall working of the proposed system

As expected by the project objective the shown Figure 4.2 displays the overall working of the project with the soil classification details displayed on the label underneath original image uploaded by the user, the outputs like model accuracy with percentage represented just next to it. All the other parameters like how much the land need water, humidity, temperature, fertilizers, suitable crops to be grown are suggested for the betterment of the farmers wealth as well as to expect more crops in future.

Step 1: Using convolutional neural networks to classify soil We process the user-provided photograph of the soil in the first stage and categories it into one of the four classificationsof soil—Red, Alluvial, Black, and Clay. This is accomplished using a convolutional neural network, the specifics of whoseimplementation are covered in the following subsections. Several crops that can be grown in that soil type areshortlisted when the soil type is forecasted. As a result,because only crops suited to the soil type are shortlisted, thisphase assures the quality of the crops advised. Training phase and testing phase were the two steps of the approach. The soil dataset and the crop dataset were both used. Chemical characteristics of soil are included in the soildataset as classes. The soil class was determined using machine learning techniques. K-NN, Gaussian Kernel based SVM.

Step 2: Yield and Income Prediction

Our algorithm takes into account characteristics such soil type (predicted by the first step), area of land to be farmed (in hectares), State, District, crop, and season of cultivation in the second phase and forecasts the yield of all crops that were shortlisted based on the parameters mentioned above (in quintals). The Random Forest Algorithm, whose implementation details are covered in the following subsections, accomplishes this.

V. CONCLUSION

A model is proposed for predicting soil series and providing suitable crop yield suggestion for that specific soil. The model will be tested by applying CNN, SVM algorithms to get the minimum required accuracy. At the end the conclusion for this project is to facilitate the farmers with the latest upgrading technologies and to let them satisfy their own need of growing multiple crops simultaneously varying with the soil type we have in different fields. We have also specified the fertilizers and the natural parameters to be maintained while producing the particular crops on the particular soil type. We have got maximum accuracy through CNN algorithm and hence the core algorithm for this model stays with CNN itself.

VI. REFERENCES

[1] Greema S Raj, Lijin Das S, “Survey On Soil Classification Using Different Techniques”, Volume: 07 Issue: 03, Mar 2020, p-ISSN: 2395-0072 International Research Journal of Engineering and Technology (IRJET).

[2] Hamzah, Mohammad Diqi, Antomy David Ronaldo, “Effective Soil Type Classification Using Convolutional Neural Network”, Vol. 3, No.1, Agustus 2021 ISSN: 2685-8711, E-ISSN: 2714-5263 DOI : 10.35842/ijicom

[3] Pallavi Srivas tava, Aasheesh Shukla, Atul Bansa, "A comprehensive review on soil classification usingdeep learning and computer vision", DOI:10.1007/s11042- 021-10544-5

[4] Shravani V, Uday Kiran, Yashaswini J, and Priyanka D, "Soil Classification And Crop Suggestion Using Machine Learning", Volume: 07 Issue: 06, June 2020, p- ISSN: 2395-0072, International Research Journal of Engineering and Technology (IRJET)

[5] S. A. Z. Rahman, K. Chandra Mitra and S. M. Mohidul Islam, "Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series," 2018 21st International Conference of Computer and Information Technology (ICCIT), 2018, pp. 1-4, doi: 10.1109/ICCITECHN.2018.8631943.