Masters of Applied Computing

Advanced Database Topics (COMP-8157)

Project Phase - 2

Title:

Soil Analysis Using OpenCV, TensorFlow and TF.Sort

Group Name:

Data Irrigators (Group 4)

Submitted to:

Dr. Shafaq Khan

Student Name	Student ID
Dhruman Rathod	110094647
Khondoker Aminuzzaman	110090822
Himanshu Rao	110090268
Pawankumar Akbari	1100917370

Soil Analysis Using

OpenCV, TensorFlow and TF.Sort

Abstract:

This paper proposes a method to categorize soil types and predict their quality for certain agricultural crops using image processing and deep neural networks. The suggested approach intends to increase sustainable agriculture and improve food security by giving farmers more accurate and useful tools for soil analysis that can increase crop output. The paper reviews several studies that have utilized image processing, machine learning, and deep neural networks for soil analysis. The proposed method uses TensorFlow sort, Matlab, Sklearn, OpenCV, and SVM algorithms for soil classification, and k-means for simplifying the classification and resulting in Feature extraction which further helps us to amplify the exact soil classification. The paper also discusses the functional requirements for the proposed method, including OpenCV and image processing. The article concludes that the proposed method can improve the accuracy of soil classification and prediction of crop quality, which can have a significant impact on sustainable agriculture and food security.

Introduction:

The productivity and sustainability of agricultural regions are largely determined by the health and fertility of the soil. Since agriculture is critical to the livelihoods of a significant portion of the world's population, it is necessary to carefully examine the soil and categorize it to match the appropriate crops with suitable soil types. However, conventional soil analysis techniques can be time-consuming, expensive, and labor-intensive. This study aims to investigate the use of image processing and neural networks to classify soil types and forecast their quality for specific agricultural crops. The integration of OpenCV, TensorFlow, and deep neural networks could help enhance food security and promote environmental conservation for future generations. The research questions focus on soil quality classification and prediction, the use of image processing to identify soil chemical composition and nutrient content, and the application of soil granules with the TF.sort algorithm to implement bifurcation based on indexing. This study's outcomes may have a significant impact on sustainable agriculture and food security by providing farmers with more precise and effective tools for soil analysis, which will improve crop output.

Literature Review:

In this section, we have summarized some papers that have worked on soil analysis using image processing, machine learning, and deep neural networks.

A combination of image processing, artificial neural network, and MATLAB was utilized to determine the pH level of soil and the presence of essential nutrients such as Calcium, Potassium, Nitrogen, Zinc, Phosphorous, and Magnesium using Soil Test Kit (STK) and Rapid Soil Test Kit (RTK). The program was able to provide accurate information, but it's precision could be improved with a more extensive soil dataset. Furthermore, the program did not have the capability to recommend fertilizers based on the test results for specific plants. [1].

Different machine learning techniques such as Bagged Trees, weighted K-Nearest Neighbor (K-NN), and Support Vector Machines (SVM) using Gaussian Kernel have been employed to categorize soil and provide recommendations for crops based on the outcomes. The research was conducted on a limited dataset, and the precision of the findings could be enhanced by incorporating additional datasets into the model in the future [2].

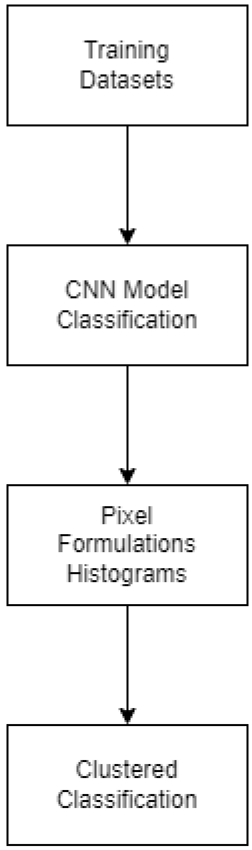
Several soil sensors, including NPK sensors, moisture sensors, and soil pH sensors, were utilized to obtain data from soil samples for classification purposes. Both Support Vector Machines (SVM) and Artificial Neural Networks (ANN) were employed, with ANN providing superior accuracy compared to SVM. However, the study revealed that the soil classifier worked well for Silty sand but was less effective for loose classy loam soil type [3].

The research employed a range of machine learning algorithms, such as decision tree bagging, naive Bayes, Support Vector Machines (SVM), logistic regression, Boosted Regression Tree (BRT), and Random Forests (RT), to classify soil as high, medium, or low fertility. The Random Forests classifier demonstrated superior performance over the others. In the future, the team plans to forecast nutrient levels based on weather conditions to provide early estimations [4].

The article discusses the use of Convolutional Neural Networks (CNN) for the classification of different types of red soil in Telangana and it is able to provide 91 percent accurate results. The CNN model is trained on a soil image dataset using pre-processing and feature extraction techniques. Softmax activation function is used to output a probability distribution of the soil type. The use of advanced technologies in agriculture can enhance crop productivity through proper fertilizer application, and with better datasets, it can provide up to 99% accuracy [5].

Proposed Model:

The method used by the authors leverages the direct input data from Fig. 1 to perform classification using the CNN model. Instead, we'll employ the Tensor Flow sort, which directly exploits picture categorization and further streamlines the distribution of the image's pixels. The image is directly sorted using Tensor Flow Sort, and the algorithm is further trained. The deep neural network employed assists in determining how distinct soil data are changed and in analyzing future picture processing. In addition, by ensuring the characteristics, the gathered data will assist in obtaining the categorization. K-means can further simplify the classification with the use of this technique. Also, this classification aids in increasing accuracy.



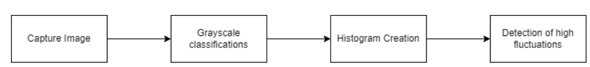
Analyzing the supplied data and making it simpler to identify the pixels is made possible by importing open computer vision libraries. Additionally, the photos are examined, taking into account each and every pixel, utilizing the pixeled matrix for histogram representation. The presence of pixel change is detected using feature detection on the histograms, and coloured images are used to identify and build color characteristics. Classifications are helped by image processing, and by using tensor flow to analyze it, the established features are employed to maintain various classes. Tensor flow assigns the details (which are primarily dependent on the attributes needed to distinguish the types of soil) to k-means and divides the images into several classes according to their classifications. SVM (Support Vector Modifier) use [8]. It is made simpler using K-means, and the final output is produced using the trained algorithm from earlier photos.

A smoothing (low pass) filter was used to remove high-frequency noise and artifacts from the image. Smoothing filters modify the value by means of a moving window operator. based on a pixel-by-pixel analysis of a small limited area of pixels. As the operator passes across the image, each pixel is affected. The smoothing filter thus gradually enhances the image by eradicating imperfections over time [7].

Methodology:

Soil analysis using OpenCV and TensorFlow can be performed using a combination of image processing techniques and machine learning algorithms. Here is a possible methodology for soil analysis using these tools:

Image Acquisition: The first step is to acquire images of soil samples using a camera or microscope. The images can be in grayscale or color and should be of high resolution and quality.



Pre-processing: The acquired images should be pre-processed to remove any noise or artifacts that may affect the analysis. This can be done using filters such as Gaussian, median, or bilateral filtering, and morphological operations such as erosion and dilation.

Segmentation: The next step is to segment the soil particles from the background using image segmentation techniques such as thresholding, edge detection, or region-based segmentation. This step helps to isolate the soil particles and make them ready for feature extraction.

Feature Extraction: Once the soil particles are segmented, features such as shape, size, texture, and color can be extracted from them using various image processing techniques such as Haralick texture features, Gabor filters, or Zernike moments.

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                Output Shape              Param #
=================================================================
conv2d_6 (Conv2D)           (None, 222, 222, 32)      896
_________________________________________________________________
batch_normalization_9 (Batch (None, 222, 222, 32)     128
_________________________________________________________________
max_pooling2d_6 (MaxPooling2 (None, 111, 111, 32)     0
_________________________________________________________________
dropout_9 (Dropout)         (None, 111, 111, 32)      0
_________________________________________________________________
conv2d_7 (Conv2D)           (None, 109, 109, 64)      18496
_________________________________________________________________
batch_normalization_10 (Batc (None, 109, 109, 64)     256
_________________________________________________________________
max_pooling2d_7 (MaxPooling2 (None, 54, 54, 64)       0
_________________________________________________________________
dropout_10 (Dropout)        (None, 54, 54, 64)        0
_________________________________________________________________
flatten_3 (Flatten)         (None, 186624)            0
_________________________________________________________________
dense_6 (Dense)             (None, 128)               23888000
_________________________________________________________________
batch_normalization_11 (Batc (None, 128)              512
_________________________________________________________________
dropout_11 (Dropout)        (None, 128)               0
_________________________________________________________________
dense_7 (Dense)             (None, 4)                 516
=================================================================
Total params: 23,908,804
Trainable params: 23,908,356
Non-trainable params: 448
_________________________________________________________________
```

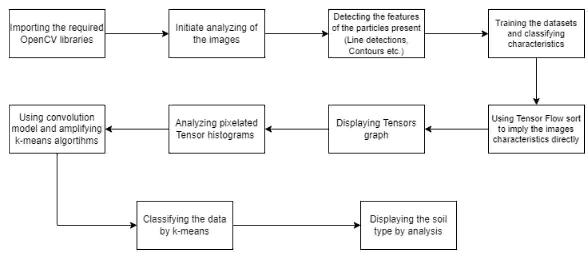
Training Data Preparation: The extracted features can be used to train a machine learning model such as TensorFlow. To do this, the soil samples should be labeled with the corresponding class or category (e.g., sand, silt, clay).

Model Training: The labeled soil samples can be used to train a machine learning model such as a convolutional neural network (CNN) using TensorFlow. The CNN can be trained using supervised learning, where the model learns to classify soil particles based on the extracted features.

Model Validation: The trained model can be validated using a separate set of test soil samples to evaluate its performance in classifying the soil particles correctly.

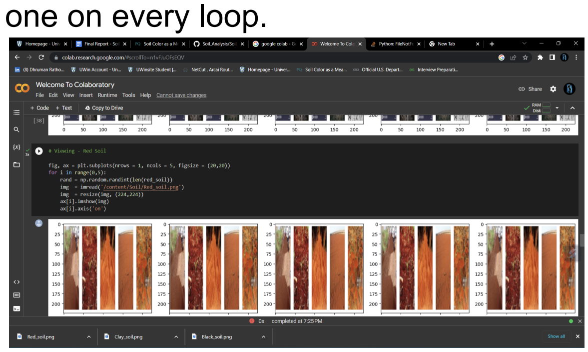
Soil Analysis: Once the model is trained and validated, it can be used to analyze new soil samples automatically. The acquired images are pre-processed and segmented, and the features are extracted from the soil particles. The trained model is then used to classify the particles into their respective classes.

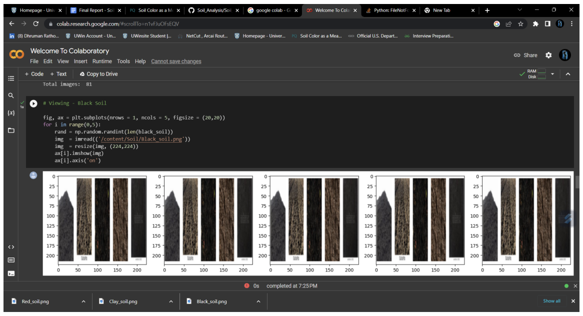
Overall, this methodology can help to automate the soil analysis process, making it faster and more accurate than traditional manual methods.



Results:

We used subplots to determine the sum of rows and columns that we showed in the plot. After that,we have used looping to display some items in the dataset, then count and columns increased by one on every loop.



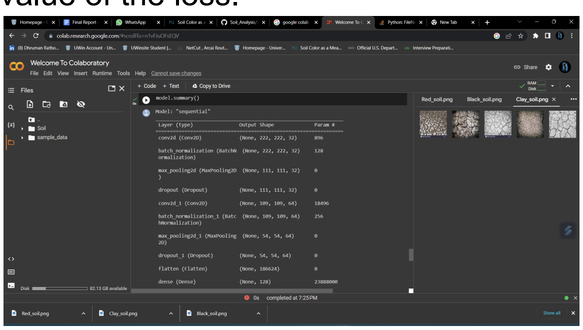


We are unable to utilize normalization since this data consists of some pictures. So, it is necessary to model this data right away using deep learning and a few layers.

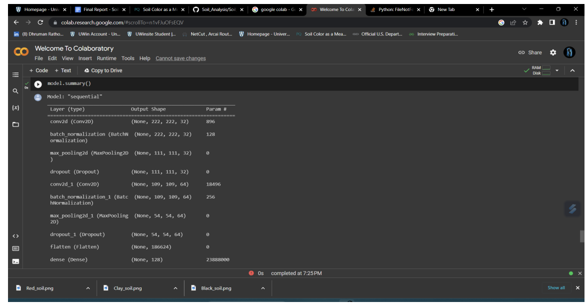
We splitted this dataset into training and testing with the sklearn library. With 80% training data and 20% testing data. And then we use the random state as much as 42 to randomize this data into 42 times with the same data so that, can't change training data and testing data.

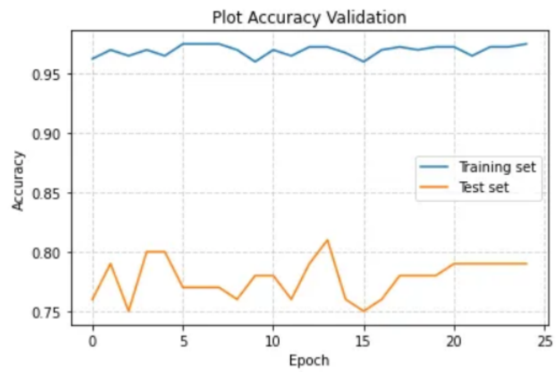
Analysis:

Based on the result accuracy from training data shown by value accuracy, whereas for accuracy validation data shown by value val_accuracy. In this case, the model was a better-learned pattern of data because it generated a high value of accuracy and a lower value of the loss.



After the modeling, we can see the final accuracy of the model prediction. About 79% of models give the true prediction of 100 testing data.





accuracy from validation data

Limitations and Challenges:

There are several limitations and challenges when it comes to analyzing soil using open cv, TensorFlow, and tf.sort. Here are some of them:

Data collection: Soil analysis requires a large amount of data, including soil samples and environmental factors such as temperature, moisture, and pH levels. Collecting this data can be time-consuming and expensive.

Data quality: The accuracy and consistency of the collected data can be a challenge. Soil properties can vary widely even within small areas, making it difficult to obtain representative samples.

Image quality: OpenCV is a computer vision library used for image processing. However, soil samples may not always provide clear and high-quality images, which can hinder accurate analysis.

Model training: TensorFlow is a popular machine learning framework used for training deep learning models. However, creating a robust and accurate model for soil analysis can be challenging, as it requires a large amount of training data and careful model design.

Interpretation of results: Once the model has been trained, interpreting the results can also be challenging. The output of a machine learning model may not be immediately interpretable by humans, and additional analysis may be required.

Limited scope: While OpenCV and TensorFlow are powerful tools, they may not be suitable for analyzing all types of soil samples or soil properties. The scope of the analysis may be limited by the available data and the capabilities of the models.

Generalization: The trained model may not be able to generalize to new, unseen data. This can be due to overfitting or biases in the training data, which can limit the applicability of the model to real-world scenarios.

Conclusion:

The ineffective methods of managing crops and the soil are the main causes of the decline in soil quality. With the aid of the OpenCV, TensorFlow, TF.Sort, the soil nutrient level is identified. Using this technique as a foundation, a neural network can be created to estimate soil fertility based on the characteristics of soil samples as seen through IoT in farming. The People can handle the land and deal with the problems associated with nutrient shortage using this approach. Additionally, this approach will be helpful in mapping the fertility of the soil and determining the fertility indices of other nutrients with a comparable composition. It may also be used in agro ecological areas for diagnosing and taking essential actions.

Future Work:

A model is suggested for determining the sort of soil and recommending a crop that can be grown in that soil. Various python libraries , including OpenCV and TensorFlow, have been used to evaluate the model. In comparison to other models, the precision of the current model is highest. In the future, appropriate fertilizers are advised for the cultivation of crops that develop well. The future models contain real-time data that is immediately received from farming land that is equipped with sensors, whereas the current models deal with outdated data that is currently accessible.The instruments detect other soil elements and the fertility of the soil [9].

References:

[1] J. C. Puno, E. Sybingco, E. Dadios, I. Valenzuela and J. Cuello, "Determination of soil nutrients and pH level using image processing and artificial neural network," in 2017 IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, Philippines, 2017, pp. 1-6. doi: 10.1109/HNICEM.2017.8269472.

[2] 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," in 2018 21st

International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 2018, pp. 1-4. doi: 10.1109/ICCITECHN.2018.8631943.

[3] J. Nanwal and P. Sethi, "Identifying Best Suited Soil Based on its Physical and Chemical Properties Using Machine Learning," in 2021 Fourth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2021, pp. 35-38. doi: 10.1109/CCICT53244.2021.00018.

[4] S. M and J. C D, "Classification of Soil Fertility using Machine Learning-based Classifier," in 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC), Jalandhar, India, 2021, pp. 138-143. doi: 10.1109/ICSCCC51823.2021.9478169.

[5] N. L. Kalyani and K. B. Prakash, “Soil Color as a measurement for estimation of fertility using Deep Learning Techniques,” International Journal of Advanced Computer Science and Applications, vol. 13, no. 5, pp. 126-131, 2022. doi: 10.14569/ijacsa.2022.0130536.

[6] Kalyani, N.L. and Prakash, K.B. (2022) “Soil Color as a measurement for estimation of fertility using Deep Learning Techniques,” International Journal of Advanced Computer Science and Applications, Vol. 13, No. 5, pp. 126-131. Available at: https://doi.org/10.14569/ijacsa.2022.0130536.

[7] S. Liu, Y. Wang, L. Li, W. Cai and C. Li, "Identification of Soil Properties Based on Image Processing Technology," Journal of Coastal Research, vol. 105, no. sp1, pp. 307-311, 2021. doi: 10.3389/fpls.2022.1092089

[8] N.Saranya and A. Mythili, “Classification of Soil and Crop Suggestion using Machine Learning Techniques,” International Journal of Engineering Research & Technology, vol. 9, no. 4, pp. 539-543, 2020.