Plano Senior High School

Determining Soil Organic Matter [SOM]

An application designed to analyze soil composition as a definitive for healthy soil

Krithika Subramanian

Computer Science / Mathematics, Engineering

Acknowledgments

This study was made possible with the encouragement of family members, teachers, and friends

Contents

1. Abstract....4
2. Introduction....5
  2.1. Proposed Solution....6
3. Background....7
  3.1 The Threat of Deteriorating Soil....7
  3.2 The Importance of Soil Organic Matter....7
  3.3 Loss-On-Ignition Analysis and Spectroscopy....7
  3.4 Analyzing Soil Color....8
  3.5 Data Mining....8
  3.6 Android Application....9
4. Materials....9
5. Procedure....9
  5.1 Repository Analysis....9
  5.2 Application....10
6. Experimentation and Results....11
  6.1 Initial Plans for Experimentation....11
  6.2 Final Form of Experimentation....11
  6.3 Data Acquisition....11
  6.4 Data Analysis....15
7. Conclusion....15
8. Error Analysis....16
9. Future Improvements for Prototype....17
10. Future Applications & Research....18
11. Works Cited....19

1. Abstract

Soil degradation, described as a decline is soil quality through its misuse by humans (aka improper agricultural techniques), is a global threat, causing increased pollution, sedimentation in drinking water sources, and crop contamination. Inappropriate agricultural farming techniques lead to soil erosion, loss of soil structure, and most importantly, a loss of biodiversity attributed to soil and land degradation.

This project aims to develop an application that can accurately and reliably allow soil samples to be tested for their composition, specifically for soil organic matter, with a smartphone camera as an alternative to expensive, less accessible spectrometer or loss-on-ignition tests. This can be done by obtaining a live camera feed of a soil sample and analyze the feed to accurately detect the average RGB color values present and determine composition by breaking down frames and isolation desired regions of interest. Specifically, the visual input would help to analyze a soil sample's organic matter, one of the most important determinants of healthy soil; a decreased amount organic matter is the earliest indicator to unhealthy soil, soil pollution, and reduced soil biodiversity. To test accuracy of the program, a database of soil samples was analyzed and tested with a simple regression model.

Future applications include implementing ensemble methods of image segmentation and neural networking improve predictions of organic matter in soil. Soil samples will also be tested with the application. The results of this project will improve existing precision agricultural technology, as well as the optical technology field as a whole.

2. Introduction

This project aims to develop an application that can accurately and reliably allow soil samples to be tested for their composition, specifically for soil organic matter, as an alternative to expensive, less accessible spectrometer or loss-on-ignition tests with a smartphone camera. Loss-on-ignition tests include burning soil to determine how much organic matter is lost, and spectrometers produce results based on analysis of soil color. However, determining soil composition solely through a camera is not only much more accessible (smartphone apps could easily serve as a consistent soil monitoring system), but also provides insight into the reliability of precision agricultural technology.

Soil degradation, described as a decline is soil quality through its misuse by humans (aka improper agricultural techniques), is a global threat, causing increased pollution, sedimentation in drinking water sources, and crop contamination. Inappropriate agricultural farming techniques lead to soil erosion, loss of soil structure, and most importantly, a loss of biodiversity attributed to soil and land degradation. The geographic extent of soil degradation is vast; it ranges from western Asia, to Central America, to various areas of Europe. In fact, in the United Kingdom, soil degradation has already been identified as a national security threat.

The purpose of this study is to develop a visual analysis technique that can be used to classify soil based on identifying characteristics like soil organic matter and other variables. By engineering a reliable, portable analysis system, determining soil composition can be vastly easier. This can be done by obtaining a live camera feed of a soil sample and analyze the feed to accurately detect the average RGB color values present and determine composition. Specifically, the visual input would help to analyze a soil sample’s organic matter, one of the most important determinants of healthy soil; a decreased amount organic matter is the earliest indicator to unhealthy soil, soil pollution, and reduced soil biodiversity.

2.1. Proposed Solution

This engineering project included the development of a mobile Android application. The proposed solution involves the usage of an algorithm to determining soil organic matter.

There are many ways this can be tested. For example, digital RGB photography and visible range spectroscopy for soil composition and analysis demonstrates the use of a spectrometer for variables like organic matter and other nutrients. Soil spectroscopy uses visible and infrared waves on the electromagnetic spectrum, which could potentially be analyzed with a camera, to identify chromophores which interact with EM radiation (clay/mineral lattices, organic matter, nonclay material, etc.). A second, non-computerized method is a Loss-On-Ignition Test, which is widely used by laboratories to determine the amount of SOM present. One method that has not been extensively researched is solely using a visual technique using a smartphone. This is the method that will be applied in experimentation. Digital precision agricultural coupled with visual analysis is a possible course of action in which the program records a live camera feed of a soil sample using an electronic device and analyzing them based on color parameters.

The finished product must be able to successfully determine the average RGB values of live camera frames, simultaneously present a running average of soil organic matter (done through linear regression), and ultimately accurately predict whether the soil is healthy.

According to the *Nature and Properties of Soil*, the concentration of SOM in soils generally ranges from 1% to 6% for most top soils. Top soils consisting of less than 1% of organic matter are limited to deserts, while wet areas can rise over 90%. Generally, top soil is classified as healthy and organic if it contains 12% to 18% soil organic matter.

3. Background

3.1 The Threat of Deteriorating Soil

A decline in soil organic matter (SOM) or soil organic carbon (SOC) is caused by the reduced presence of decaying organisms as a result of changes in natural or anthropogenic factors. SOM is a vital component of healthy soil, and its decline results in degradation. According to the European Soil Data Center (ESDAC) sources of soil organic matter include crop residue, animal/green manure, compost, or simple to complex carbon substances. In 2019, ESDAC released a study stating that organic matter in soil was on the decline, mostly due to inappropriate agricultural practices. In the UK, soil degradation has already been identified as a national security threat.

3.2 The Importance of Soil Organic Matter

SOM contributes to biodiversity as a reservoir of nutrients (nitrogen, phosphorus, sulfur, etc.), absorbs water (liquid)— a lifeline for vegetation in dry soil— and contains two times the amount of carbon in the atmosphere and three times the amount as in vegetation. When soil organic matter is formed, it takes in atmospheric carbon dioxide, thus reducing air pollution. However, when SOM is decayed, it releases carbon dioxide into the atmosphere.

The European Green Deal, published by the European Commission, presents a new soil strategy to "limit drainage of wetlands and organic soils to restore managed and dry peatlands." However, the soil strategy present establishes goals to be achieved in the future, rather than actions to be taken. Although the commission has put forth concerns about biodiversity crises, no methods to analyze soil composition have been implemented.

3.3 Loss-On-Ignition Analysis and Spectroscopy

To get an idea on the negative effect of farm management practices on the reduction in soil organic matter, soil samples should be taken over time. Most laboratories do loss-on-ignition (LOI) tests to determine the amount of SOM present over a long period of time. LOI tests are a widely-used method for measuring organic matter in soil, but have no universal standard protocol. In fact, a large number of factors can influence its accuracy: furnace type, sample mass, duration & temperature of ignition, clay content of samples, etc. The effects of ignition have scarcely been tested. Soil spectroscopy, which uses the interactions between matter and visible electromagnetic waves to determine soil composition, is inconsistent on a large scale.

The soil science community is facing a growing demand of regional, continental, and worldwide databases to monitor soil status. Using LOI or spectroscopy techniques are inconsistent and unreliable on a large scale.

3.4 Analyzing Soil Color

Soil comes in different shades with different ecosystems. Most top soils, where the most organic matter is found, are black, brown, red, gray or white. Organic matter decomposes into black humus. The different elements present in soil are often associated with colors:

Iron forms reddish-yellow crystals, manganese forms black mineral deposits, vanadium is green-purple, chromium is blue-green, nickel is blue-green or purple, copper is blue-green, zinc is pale blue, and lead is yellow.

Many elements could display a range of colors in soil depending on pH, anaerobicity, and environmental type. Soil is often described using the Munsell color system, a color space based on hue (basic color), chroma (color intensity), and value (lightness). As SOM increases, hue and chroma also increase.

3.5 Data Mining

The National Soils Inventory for Scotland (NSIS) is an objective sample of Scottish soils. The inventory stores soil and site conditions of hundreds of locations, and analyzes samples taken at multiple depths from soil pits to determine physical and chemical properties. The NSIS database is comprised of several component tables which must be manipulated and integrated into a new dataset to extract useful data.

The NSIS database will essentially be used to data mine; that is extract useful data and discover patterns in a large dataset through machine learning and statistical analysis. Log transformations will be applied to make highly skewed distributions less skewed and make the data more interpretable. The attributes and relations of the useful data will also be normalized for further accuracy.

Linear regression will be used to model the relationship between RGB color values and SOM in percent. The loss function of regression will be the Mean Squared Error (MSE) to optimize the model. Other models, such as neural networking, may also be implemented in the future.

$$Y=\beta_1 x_1+\beta_2 x_2+\beta_2 x_2+b$$

Where Y is the SOM (%/100); β_x are the coefficients or unknown model parameters estimated from the data that changes with each run-through; x_x are the three inputs of RGB values (1. Red 2. Green 3. Blue); and b is the intercept.

3.6 Android Application

The user interface for developing the Android application followed basic material design guideline. The Android application was tested on a Samsung 2015 J7 smartphone running API 26 (Android 8.0.0. Oreo).

The Java IDE used to program the application was Android Studio.

4. Materials

Minimum hardware requirements:

- Any Android device running version 8.0.0 (Android Oreo, API 26) or later with a functional camera
- A laptop running Windows 10 version 1903 or later with at least 2 GB RAM and 1.5 GB available hard disk space

Minimum software requirements:

- Any Java IDE compatible with Android software, such as Android Studio, Eclipse, or IntelliJ IDEA

Soil samples or repository of data

5. Procedure

5.1 Repository Analysis

The National Soils Inventory of Scotland was used in testing the program, and its data was analyzed using a Python IDE (PyCharm, IntelliJ IDEA, etc.)

1. Load dataset from the National Soils Inventory of Scotland, or other compiled repository
    a. Convert the files into a readable format (shape - .shp files into readable comma separated - .csv files) and load it into memory
    b. Review relevant columns of data (Soil Organic Matter, nutrients, etc.)
2. Modify the data: convert Munsell color values into an RGB format
    a. Convert data into properly formatted Munsell color values in code
    b. Convert Munsell values into RGB color system
    c. Output (Munsell form, [Red, Green, Blue])
    d. Apply conversion to specific columns and remove any unnecessary or unusable data
3. Perform log transformations on all graphs to remove skewed data and zoom in on areas of interest
4. Perform normalization and zoom in on areas of interest for further analysis
5. Create a correlation matrix between colors and composition
6. Model and create predictions: the features are the three color parameters (R, G, B) and the target is amount of organic matter present in percentage. This is a supervised learning machine, since a target is given. This is a regression task, because the target is a continuous variable.
    a. The first type of model tested will be a linear regression model, with mean squared error as the loss function, minimized using gradient descent.
    b. The data is separated into training and testing sets (70/30 train to test split)
        i. There is a total of 2701 items, so 1890 items were trained and 811 tested

5.2 Application

1. Develop an application using a Java IDE that can determine the soil composition and percentage of organic matter in a soil sample for use on an Android device. This application will obtain data from the device's camera.
2. Produce a live feed of the soil sample (data from soil inventory databases can be analyzed for further research and predictions).
3. The live camera feed will be broken down to isolate the desired region of interest, allowing us to determine the percentage of inorganic and organic matter.

4. Color parameters of the image can be used to estimate the chemical composition of the soil.
5. The application should accurately determine the sample's sand, silt, and clay percentages, as well as organic and inorganic matter percentages.

6. Experimentation and Results

6.1 Initial Plans for Experimentation

Originally, the visual input was going to be compared with a repository in the application using a Fast Fourier Transform implementation, however it was deemed impossible because the repository used for data mining (NSIS) consists of Scottish soils, which would not be consistent with soil samples in the United States.

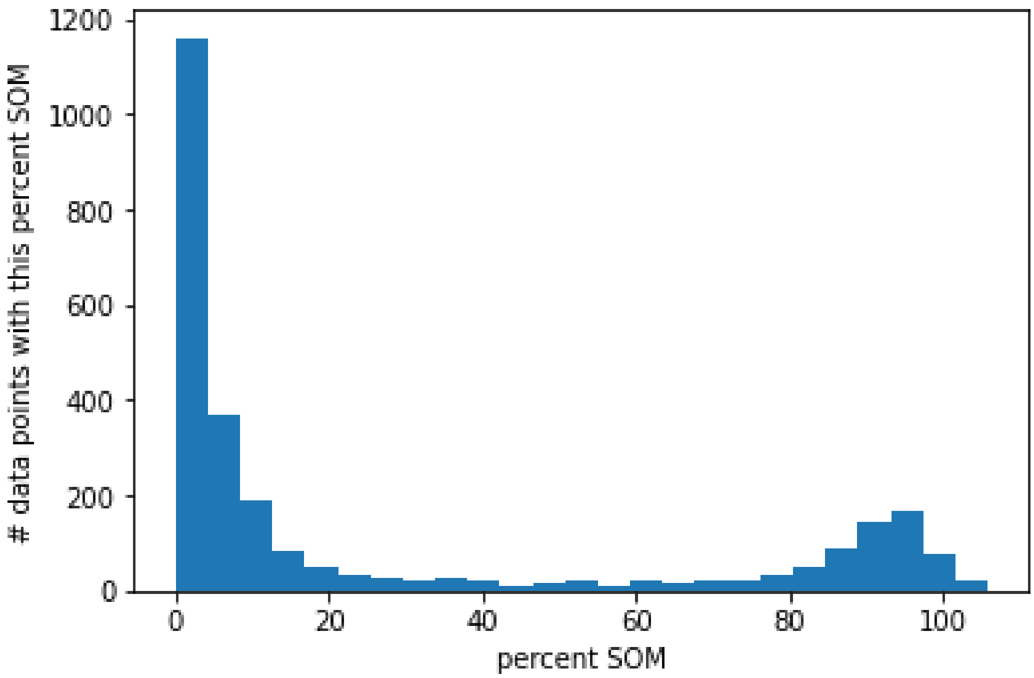
6.2 Final Form of Experimentation

Instead of using a repository, a live camera feed was analyzed, and the average RGB values were inputted into a simple linear regression technique to determine the amount of SOM present.

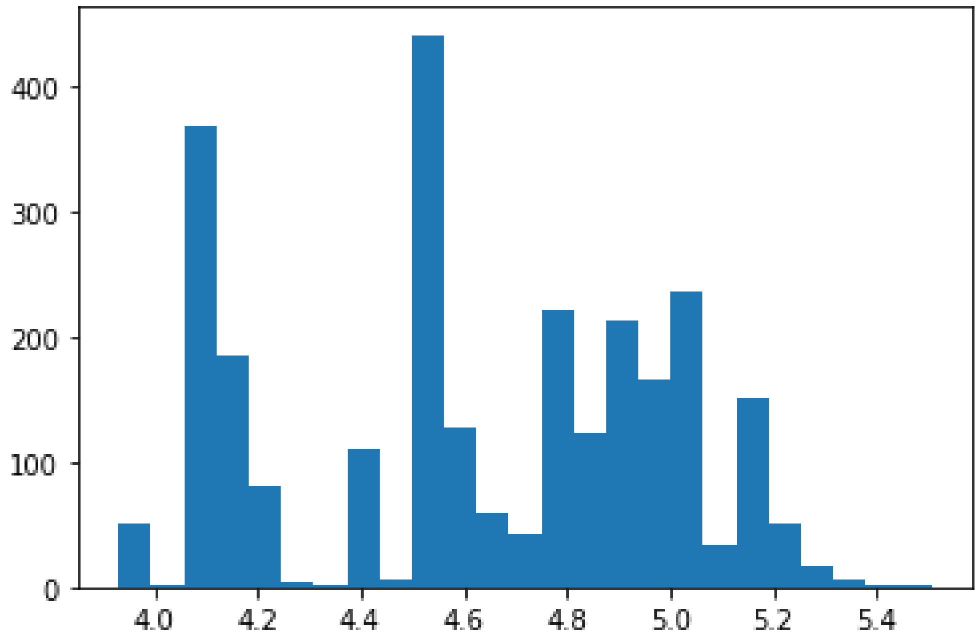
6.3 Data Acquisition

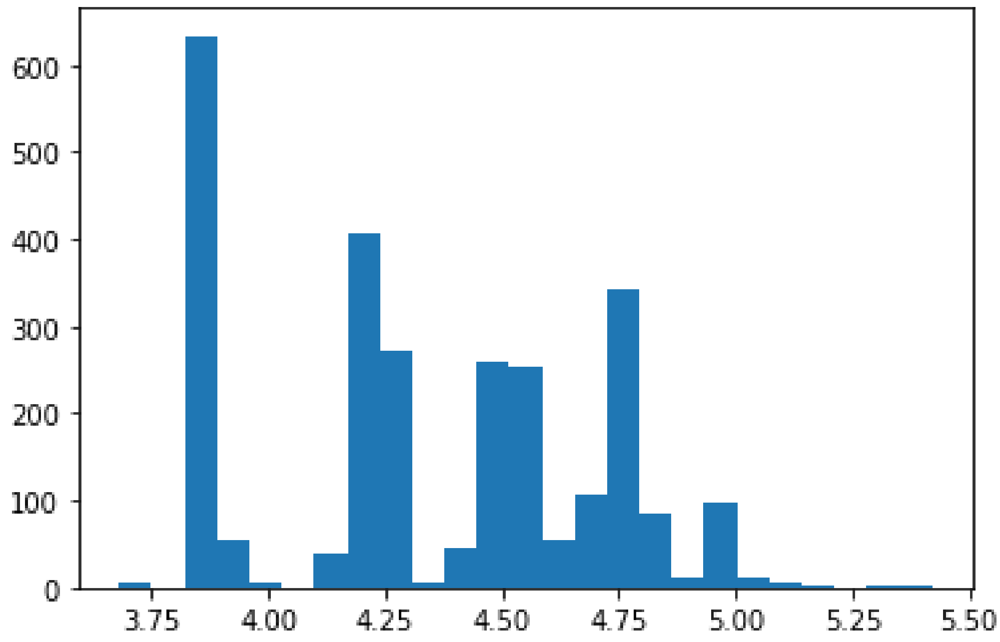
The following data consists of NSIS repository analysis. Implementations for the application are low accuracy prototypes and will be improved in the future.

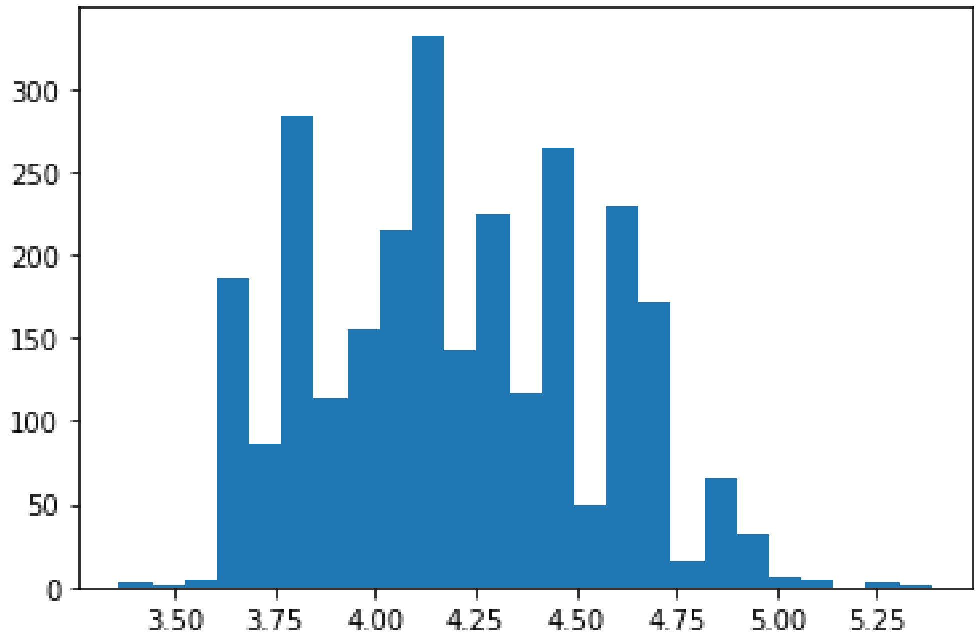
<u>Analysis of All Data Points</u>


1200
1000
800
600
400
200
0
data points with this percent SOM
0
20
40
60
80
100
percent SOM

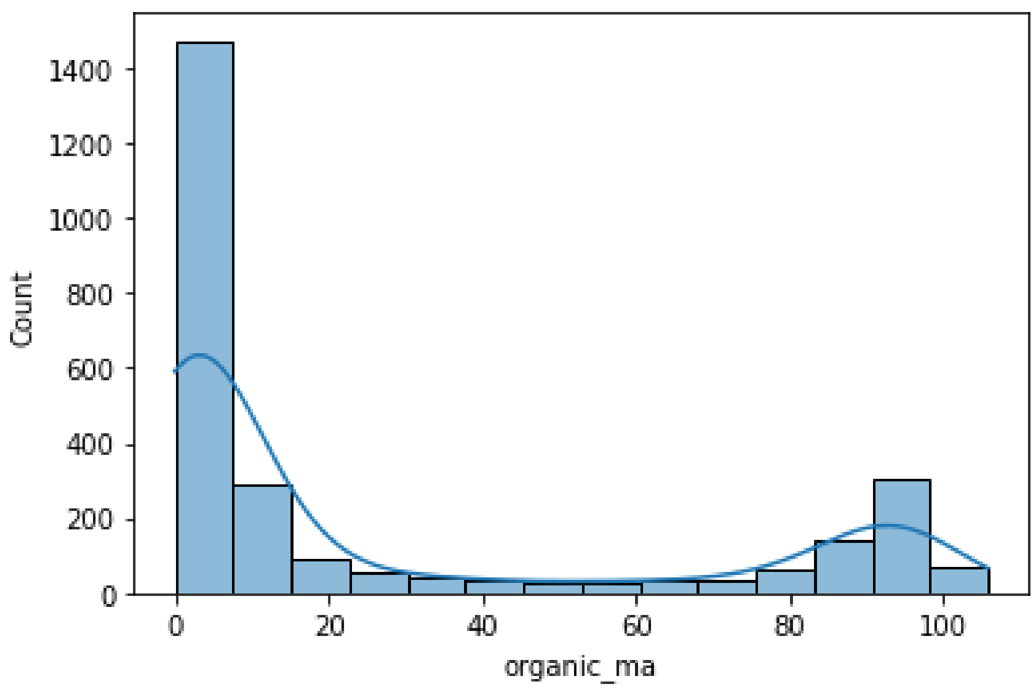
Log Transformations



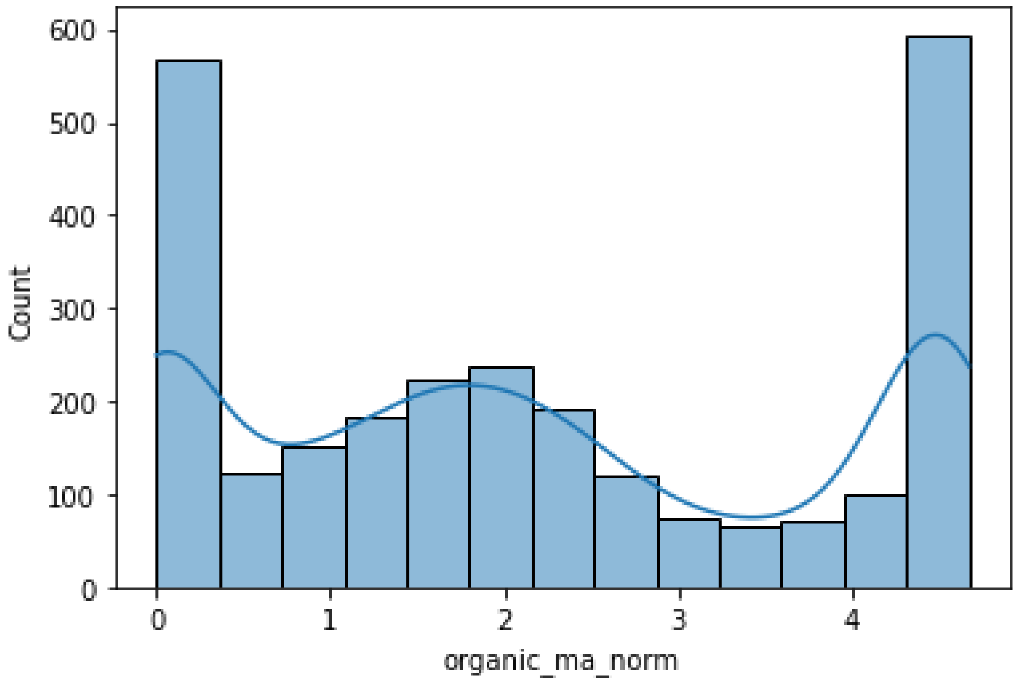




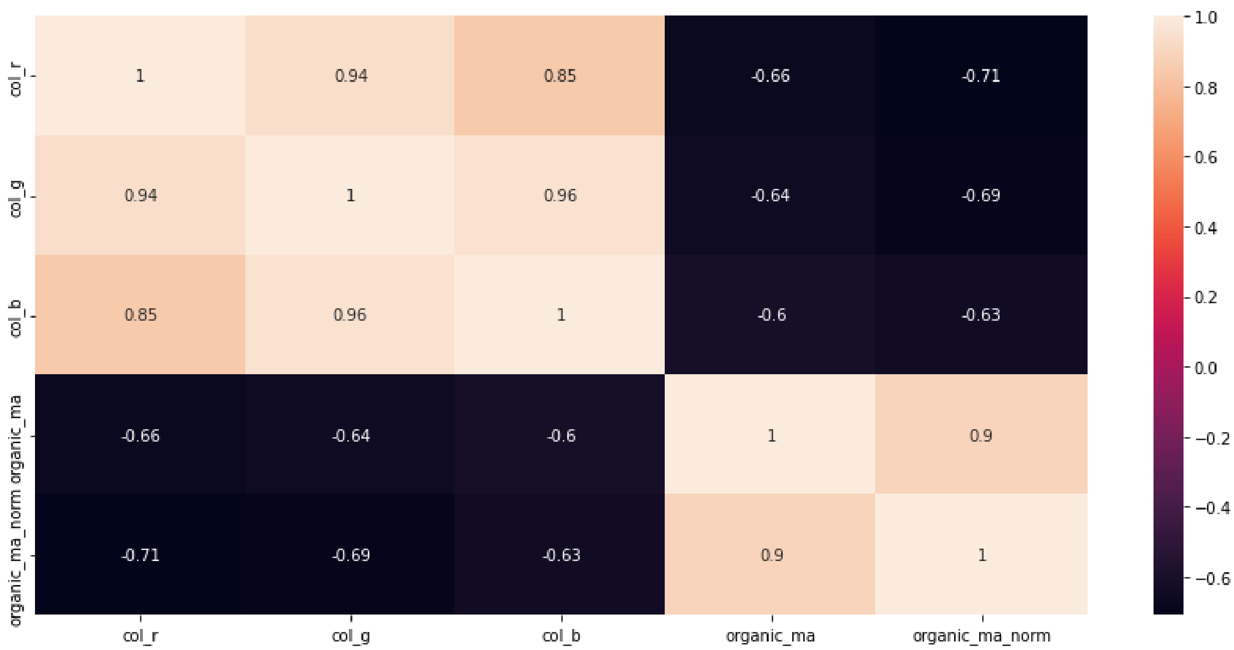
After Normalization, Full View



After Normalization, Zoom-In



Color Correlation Matrix



Model Predictions

(4 random soil samples from repository shown below)

	Red value	Green value	Blue value	Predicted SOM (%)	Actual SOM (%)	Accuracy (%)
01	156	97	70	2.66	1.7	56.5
02	66	44	30	58.67	82.6	29.0
03	102	7	6	65.96	88.8	25.7
04	66	99	30	5.25	4.0	31.3

6.4 Data Analysis

Through data mining, an analysis of all data points were found. It was shown that the data was skewed highly to the left, so log transformations were taken and zoomed-in on. To further improve accuracy, normalization was done, however it did not work as intended, as most points were still skewed left towards 0%.

A developed color correlation matrix shows a clear trend between colors and SOM. The darker the color (the higher the RGB values are), the higher the SOM percentage will be.

After data extraction, the final training set consisted of 1890 items, while the testing set consisted of 811 items. Four random data points were chosen and tested with the linear regression program, resulting in an average accuracy of 35.1%.

7. Conclusion

The current soil degradation crisis poses an extreme threat to our world, and ultimately, our future. In the United Kingdom, soil degradation has already been identified as a national threat, and the soil science community is facing a growing demand of regional, continental, and worldwide databases to monitor soil status. This is largely due to inaccessible and expensive composition tests, like loss-on-ignition testing or soil spectroscopy, which can not only cost upwards of one-thousand USD, but are also not available in many smaller or local laboratories. The soil composition tests mentioned utilize electromagnetic waves and the amount of mass loss when burned to determine the amount of SOM present. However, detecting the likelihood of healthy soil solely through visual techniques is not only much more accessible (smartphone apps could easily serve as a soil monitoring system over a long period of time), but also provides insight into the reliability of precision agriculture technology. The purpose of this study is to develop a optical analysis technique that can be used to classify soil samples based on their SOM percentage. This can be done by obtaining the average RGB value for each frame of a live camera feed and inputting into a simple linear regression equation. The National Soils Inventory of Scotland was used to test the machine learning program, and through data mining, promising results were found.

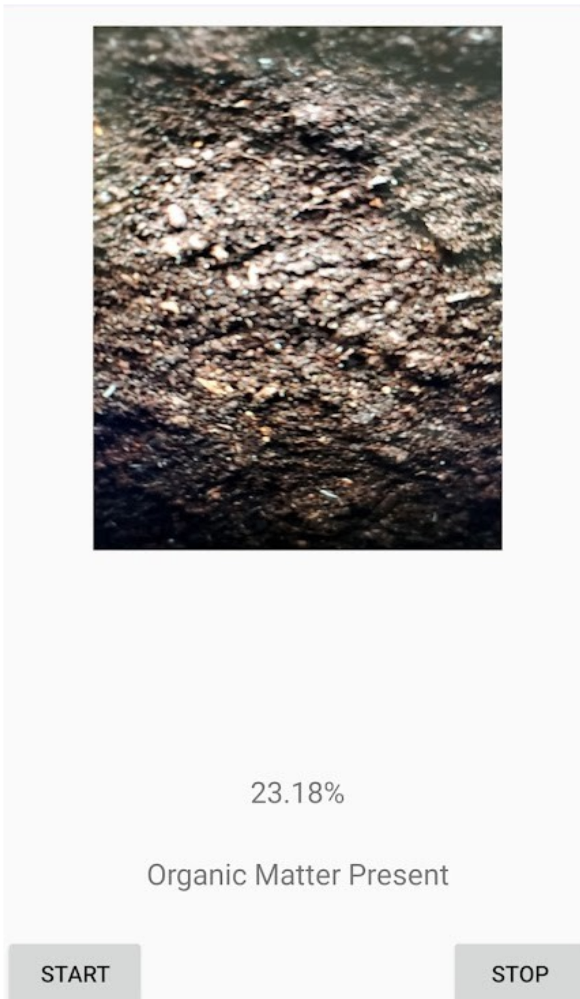
The results showed that the developed software was able to determine the amount of soil organic matter present with a training set comprised of 1890 items and a testing set comprised of 811 items (a 70/30 split). It was found that the training set had 43.8% accuracy, while the testing set had 44.9% accuracy. While these percentages seem low, it is to be expected since only a simple regression technique was used. In the future ensemble methods of image segmentation and neural networking will be used to improve predictions of organic matter in soil.

8. Error Analysis

Unfortunately, the results of this experiment may not have been accurate because only the NSIS database was tested with the program, and the regression technique used could be improved. To view clear results, more testing needs to be done with the live camera feed, with various soil samples, and a different machine learning model, like neural networking. Ultimately, the data collected satisfied the goal of this project because the developed software was able to carry out a digital precision agriculture task quickly and somewhat reliably with each participant tested. Furthermore, the results of this experiment can be used to expand upon the reliability and use of precision technology, especially for soil-monitoring purposes.

9. Future Improvements for Prototype

Some basic improvements to the prototype are to test different models, such as neural networking, produce a higher accuracy. Preprocessing techniques like image segmentation can also be implemented for higher accuracy. The current linear regression technique produces a low percent accuracy, so more testing needs to be done to improve the program. A prototype of the application is shown below.



10. Future Applications & Research

Future applications of the prototype are to take into account the type of ecosystem where the soil is found (for example, forest, grassland, desert, etc.). In addition, the application should be able to exhibit an intensive analysis of soil composition (sand, silt, clay, nutrients, etc.), rather than just analyze soil organic matter. In the future, a database for soils in a specific region can be established so results can be saved over a long period of time and consistency of soil matter can be proven.

In the future, the results of this project can be used to improve the existing software for usage in various tracts of land and environments. The results of this project can also be used to improve and impact existing precision agricultural and visual analysis technology, as well as the optical technology field as a whole. Furthermore, this application can be tested on different environments, particularly dry soil samples compared to wet soil samples, and can be tested in conjunction with LOI testing or spectroscopy to determine if all tests produce similar results.

11. Works Cited

Aitkenhead, M., Cameron, C., Choisy, B., Coull, M., & Black, H.(2018). Digital RGB photography and visible-rangespectroscopy for soil composition analysis. Geoderma, 313,265e275.

Aitkenhead, M. J., Coull, M., Towers, W., Hudson, G., &Black, H. I. J. (2013). Prediction of soil characteristics andcolour using data from the National Soils Inventory ofScotland. Geoderma, 200e201, 99e107. https://doi.org/10.1016/j.geoderma.2013.02.013

Brady, N. C., and Weil, R. R. *The Nature and Properties of Soils.* Prentice Hall, Inc., Upper Saddle River, New Jersey, USA, 1999.

Cleland, Thomas M. (1921). *A practical description of the Munsell color system, with suggestions for its use.* Boston: Munsell Color Company.

Department of Crop and Soil Sciences. (2008). *Soil Organic Matter Agronomy Fact Sheet Series*. http://franklin.cce.cornell.edu/resources/soil-organic-matter-fact-sheet

Doi, R., Ranamukhaarachchi, S.L., 2007. Soil colour designation using Adobe Photoshop™in estimating soil fertility restoration by Acacia auriculiformis plantation ondegraded land. Current Science 92 (11), 1604–1609

Dudley, R.J., 1975. The use of colour in the discrimination between soils. Journal of theForensic Science Society 15 (3), 209–218

European Commission. (2021, November 17). *EU soil strategy for 2030*. Ec.europa.eu. https://ec.europa.eu/environment/publications/eu-soil-strategy-2030_en

Gorthi, Srikanth & R K, Swetha & Chakraborty, Somsubhra & Li, Bin & Dutta, Sudarshan & Banerjee, Hirak & Das, Krishnendu & Majumdar, Kaushik. (2021). Soil organic matter prediction using smartphone-captured digital images: Use of reflectance image and image perturbation. Biosystems Engineering. 209. 154-169. 10.1016/j.biosystemseng.2021.06.018.

Hoogsteen, M. J. J., Lantinga, E. A., Bakker, E. J., & Tittonell, P. A. (2018). An Evaluation of the Loss-on-Ignition Method for Determining the Soil Organic Matter Content of Calcareous Soils. *Communications in Soil Science and Plant Analysis*, *49*(13), 1541–1552. https://doi.org/10.1080/00103624.2018.1474475

SoCo Fact Sheets - ESDAC - European Commission. (n.d.). Esdac.jrc.ec.europa.eu. Retrieved November 30, 2021, from https://esdac.jrc.ec.europa.eu/projects/soco-fact-sheets

Soil Color | NRCS Wisconsin. (2019). Usda.gov. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/wi/soils/?cid=NRCSEPRD1370419

Moritsuka, N., Matsuoka, K., Katsura, K., Sano, S., & Yanai, J. (2014). Soil color analysis for statistically estimating total carbon, total nitrogen and active iron contents in Japanese agricultural soils. *Soil Science and Plant Nutrition*, *60*(4), 475–485. https://doi.org/10.1080/00380768.2014.906295

Owens, P. (2014). *Soil Color - an overview | ScienceDirect Topics*. Sciencedirect.com. https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/soil-color

Sánchez-Marañón, M. (2011). Color Indices, Relationship with Soil Characteristics. *Encyclopedia of Agrophysics*, 141–145. https://doi.org/10.1007/978-90-481-3585-1_237