Soil Classification using machine learning models

Thesis submitted in partial fulfillment for the award of the degree

of

B.Tech in Electronics & Communication Engineering

Submitted By

Harsh Parashar (B21EC033)

Koninika Tarafdar (B21EC034)

Under the guidance of

Dr. Salam Shuleenda Devi

(Asst. Prof. of Electronics & Communication Engineering)

DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY MEGHALAYA

LAITUMKHRAH, BIJNI COMPLEX, SHILLONG-793003

6TH MAY 2024

राष्ट्रीय प्रौद्योगिकी संस्थान मेघालय

NATIONAL INSTITUTE OF TECHNOLOGY MEGHALAYA

Bijni Complex, Laitumkhrah, Shillong 793003

ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE OF APPROVAL

The project titled **"Soil Classification using machine learning models"** is hereby approved as a creditable study of engineering subjects carried out by **Harsh Parashar (B21EC033)** and **Koninika Tarafdar (B21EC034)** and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood by this approval that the undersigned do not endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve only for the purpose for which it has been submitted.

Dr. Salam Shuleenda Devi
Assistant Professor, ECE
NIT Meghalaya

Dr. Prabir Kumar Saha
Head of the Department
Electronics and Communication Engineering
NIT Meghalaya

CANDIDATES DECLARATION

We declare that this project titled **"Soil Classification using machine learning models"** prerequisite towards partial fulfillment for the award of degree in Electronics & Communication Engineering submitted to the Department of Electronics & Communication Engineering, NIT Meghalaya is an accurate record of our work carried out under the guidance of **Dr. Salam Shuleenda Devi**, Electronics & Communication Engineering Dept. NIT Meghalaya.

Harsh Parashar
(B21EC033)

Koninika Tarafdar
(B21EC034)

ACKNOWLEDGEMENTS

We would like to express our deep gratitude and sincere thanks to our project guide, Dr. Salam Shuleenda Devi of Electronics and Communication Engineering, NIT Meghalaya for her guidance and active support during the progress of our project. Without her support and encouragement this project would have been trivial.

We also extend our heartiest thanks to our project coordinator Dr. Satyendra Singh Yadav and Dr. Shravan Kumar Bandari, Faculties of Electronics and Communication Engineering Department for their cooperation and guidance.

We would also like to mention that it would not have been possible without the timely help and support of Electronic and Communication Engineering Department Laboratories and the technical staff working there.

On a more personal note, we would like to express our heartiest thanks to our parents, all our seniors and friends, who directly or indirectly, contributed to the successful completion of our project.

ABSTRACT

This project focuses on the classification of soil data images using machine learning algorithms based on various texture and color features. Leveraging texture analysis techniques such as GLCM, wavelet transform, and mask convolution, alongside color feature extraction methods like RGB, HSV, CMYK, and color moment, the aim is to discern distinct patterns and characteristics in soil samples. Through rigorous experimentation and analysis, it evaluates the performance of prominent machine learning algorithms including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Neural Network (NN) classifier, and Kernel classifier across different feature combinations. Our findings underscore the significance of feature selection and algorithm choice in achieving optimal classification accuracy. Additionally, In the future scope of this project, envisioning advancements in soil analysis using deep learning methodologies for nutrient analysis. By expanding the dataset of soil images and developing a web application for soil classification and nutrient prediction, and strive to democratize access to advanced soil analysis tools and foster sustainable agricultural practices.

Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Introduction to the soil

Soil is the dynamic, living skin of the Earth, comprising a complex mixture of minerals, organic matter, water, air, and organisms. It forms through the slow process of weathering, as rocks break down and organic materials decompose over time[1]. Soil is not a static entity but a constantly evolving system influenced by geological, climatic, biological, and human factors. Its composition and properties, including texture, structure, fertility, and pH, vary widely across different regions and landscapes.

The significance of soil lies in its critical role in sustaining life on Earth. It serves as a medium for plant growth, providing anchorage, nutrients, and water essential for vegetation. Moreover, soil hosts a myriad of organisms—from microscopic bacteria to earthworms—participating in nutrient cycling, decomposition, and soil formation processes. Soil also acts as a natural filter, purifying water as it percolates through its layers.

Human activities, however, pose significant threats to soil health and functionality. Practices such as deforestation, intensive agriculture, urbanization, and industrial pollution can degrade soil quality, leading to erosion, nutrient depletion, compaction, and loss of biodiversity.

Given its fundamental importance, the conservation and sustainable management of soil are paramount. Protecting soil health not only ensures food security, water quality, and biodiversity conservation but also mitigates climate change by sequestering carbon dioxide. Preserving this precious resource is essential for the well-being of both current and future generations.

1.2 Formation of the soil

Soil formation is a complex and gradual process driven by various factors. It typically begins with the weathering of parent material, such as rocks and minerals, through physical, chemical, and biological processes. Physical weathering involves the mechanical breakdown of rocks into smaller particles due to factors like temperature fluctuations, frost action, and the growth of plant roots. Chemical weathering occurs as minerals within the rocks react with water, acids, and gases, leading to their decomposition and alteration. Biological activity, including the actions of plants, microorganisms, and burrowing animals, further contributes to soil formation by breaking down organic matter and enhancing nutrient cycling.

Over time, these processes lead to the accumulation of organic matter, minerals, and nutrients, gradually developing distinct soil horizons or layers. Climate, topography, parent material, and time all influence the rate and nature of soil formation[2]. For instance, climates with high rainfall and temperatures generally experience faster weathering rates, leading to the development of thicker soils.

Through this intricate interplay of physical, chemical, and biological processes, soil evolves into a diverse and fertile medium essential for supporting terrestrial life. Understanding soil formation processes is crucial for sustainable land management practices, agriculture, and environmental conservation efforts.

1.3 Types of soil

Type of soil varies according to the region or climate of its location.Each layer depth will have different levels of nutrients and minerals, all of which are very important. Some of these soil types are very similar, but all have their unique characteristics and physical properties. As soil formation happens over time, that is what forms the layers and distributes certain minerals, such as nitrogen and phosphorus, throughout the soil. Learning the soil type will determine the best use of the soil.

Soil can be classified on the basis of texture features, color features, and composition. Some of the most widely available and useful soil types are - sandy soil, clay soil, silt soil, loamy soil, peaty soil, chalky soil, alluvial soil, black soil, laterite soil, sandy soil, cinder soil, yellow soil.

(a) Black Soil

(b) Cinder Soil

(c) Laterite Soil

(d) Peat Soil

(e) Yellow Soil

Fig 1.1: Types of soil

1.3.1 Black soil

Black soil, also known as black cotton soil or regur soil, is a type of fertile soil characterized by its dark color and high organic matter content as shown in Fig1.1(a). Found in regions with a semi-arid to sub-humid climate, such as the Deccan Plateau in India, black soil is renowned for its fertility and ability to support a wide range of crops. Its unique composition, rich in clay minerals like montmorillonite, imparts excellent moisture retention and nutrient-holding capacity. Black soil is highly valued in agriculture for its ability to sustain crop growth even during dry spells. Despite its fertility, black soil can become compacted and prone to waterlogging if not managed properly.

Black soil is renowned for its exceptional nutrient content, making it highly suitable for agricultural cultivation. It is rich in essential nutrients such as nitrogen, phosphorus, and potassium, crucial for plant growth and development. Additionally, black soil contains abundant micronutrients like iron, manganese, and zinc, which play vital roles in various metabolic processes within plants. The high organic matter content in black soil contributes to its nutrient-retention capacity and promotes soil fertility. Proper management practices, including crop rotation, organic amendments, and balanced fertilization, help maintain and enhance the nutrient levels in black soil, ensuring sustained agricultural productivity and soil health

1.3.2 Cinder soil

Cinder soil, also referred to as volcanic soil or volcanic ash soil, is a unique type of soil formed from the weathering and decomposition of volcanic rocks and ash as shown in Fig1.1(b). Typically found in regions with a history of volcanic activity, such as around volcanic cones or in volcanic islands, cinder soil exhibits distinct characteristics shaped by its volcanic origins. The texture of cinder soil can vary widely depending on factors such as the type of volcanic material, the degree of weathering, and the presence of other minerals and organic matter. However, common traits of cinder soil include a coarse texture, porous structure, and light coloration. Due to its porous nature, cinder soil often has excellent drainage properties, making it suitable for cultivation in areas prone to heavy rainfall or irrigation

Analyzing the nutrient content of cinder soil reveals its potential for agricultural productivity and soil health. While cinder soil may lack organic matter compared to other soil types, it often contains a rich array of minerals and nutrients derived from volcanic material. Key nutrients found in cinder soil include potassium, phosphorus, and calcium, which are essential for plant growth and development. These nutrients are released gradually as the volcanic rocks weather over time, contributing to the long-term fertility of cinder soil. Furthermore, cinder soil may contain trace elements such as magnesium, sulfur, and boron, which play crucial roles in various biochemical processes within plants. However, the nutrient composition of cinder soil can vary depending on factors such as the age of the volcanic material, the surrounding environment, and the presence of other minerals. To optimize agricultural productivity in cinder soil, farmers may employ strategies such as soil amendments, crop selection, and irrigation management to address specific nutrient deficiencies and enhance overall soil fertility.

1.3.3 Laterite soil

Laterite soil, characterized by its red color due to iron oxide content, is commonly found in tropical and subtropical regions with high temperatures and heavy rainfall as shown in Fig1.1(c). Formed through the weathering of rocks rich in iron and aluminum minerals, laterite soil exhibits unique properties that influence agricultural practices. Despite its low fertility and poor nutrient content, laterite soil can be agriculturally productive with proper management techniques. Its porous nature allows for good drainage, reducing the risk of waterlogging. However, laterite soil often requires significant soil amendments and management strategies, such as adding organic matter and using appropriate fertilizers, to improve fertility and support crop growth effectively

Laterite soil typically exhibits low fertility and nutrient content, posing challenges for agricultural productivity. It is often deficient in essential nutrients such as nitrogen, phosphorus, and potassium, limiting plant growth and development. Additionally, micronutrient deficiencies, particularly in iron and zinc, are common in laterite soil, further impacting crop yields. Despite its nutrient limitations, laterite soil contains high levels of iron and aluminum oxides, contributing to its characteristic red color. To address nutrient deficiencies, farmers may employ strategies such as soil amendment with organic matter, application of fertilizers tailored to specific crop needs, and crop rotation to manage nutrient uptake. Effective nutrient management is essential for maximizing agricultural productivity and sustainability in laterite soil regions.

1.3.4 Peat soil

Peat soil is characterized by its high organic matter content, primarily composed of decomposed plant material accumulated over thousands of years in waterlogged environments as shown in Fig1.1(d). It has a dark brown to black color and a spongy texture, retaining moisture exceptionally well. Peat soil is acidic and often found in wetland areas or regions with high rainfall. While it offers excellent moisture retention, it can become waterlogged if drainage is inadequate. Despite its challenges, peat soil is prized for its fertility and ability to support unique plant communities, including acid-loving species like rhododendrons and blueberries.

Peat soil contains rich stores of organic matter, which slowly release nutrients as they decompose. It typically harbors high levels of nitrogen, phosphorus, and potassium, essential for plant growth. However, peaty soils may have low levels of other nutrients like calcium and magnesium, requiring supplementation for optimal plant development. Gardeners and farmers often amend peaty soil with lime to reduce acidity and balance nutrient availability. While challenging to manage due to its tendency to compact and become waterlogged, peaty soil can be highly productive when properly cultivated and managed

1.3.5 Yellow soil

Yellow soil, also known as yellow earth or yellow-brown soil, is a type of soil characterized by its distinctive color and composition as shown in Fig1.1(e). This soil is typically found in regions with a temperate to subtropical climate, such as parts of Asia, Europe, and North America. Yellow soil derives its color from the presence of iron oxides, particularly hematite and goethite, which impart a yellowish-brown hue to the soil profile. The formation of yellow soil is influenced by various factors, including the weathering of parent rock material, leaching processes, and the accumulation of organic matter over time. As a result, yellow soil can exhibit diverse textures ranging from sandy to loamy, with variations in fertility and drainage properties.

Analyzing the nutrient content of yellow soil provides insights into its suitability for agricultural use and ecosystem functioning. While yellow soil may not possess the high organic matter content of other soil types like black soil, it often contains a balanced array of essential nutrients for plant growth. Key nutrients present in yellow soil include nitrogen, phosphorus, and potassium, which are crucial for supporting plant metabolism and development. These nutrients are replenished through natural processes such as weathering of minerals in parent material and organic matter decomposition. Additionally, yellow soil may contain micronutrients such as manganese, copper, and zinc, which play vital roles in enzymatic reactions and nutrient uptake by plants.

1.4 Literature Review

Table 1.1: Literature Review

Author(s) Year	Goal	Classifier	Feature Sets	Accuracy(%)
Pathkar s, et al, 2018 [3]	Determining the texture of image using statistical and geometrical features	SVM	Color moment, HSV, wavelet transform and gabor filter	87.5
Shenbagavalli ., et al 2011 [5]	Utilizing Law's mask convolution to extract features from images to quantify texture content	1D convolutional kernels	Law's mask to gain 5 statical texture parameters	99.3
Maniyath SR, et al 2028 [7]	Detect soil color using image processing	K-NN	HSV, Median filter, RGB index, CMYK	89.3
Barman U et al 2020 [8]	Soil texture classification using multi-class support vector machine	SVM	HSV histogram color moments, and discrete wavelet transform	91.37
Shukla G., et al, 2018 [9]	Random Forest classifier as a soil spatial predictive model	Random Forest	'SCORPAN' parameters	91.2

1.5 Problem Finding

Soil classification is vital for nutrient analysis due to its direct impact on nutrient availability. Different soil types possess varying compositions and properties affecting nutrient retention, pH levels, texture, structure. Clay soils retain more nutrients, while sandy soils leach them faster. pH influences nutrient availability; Soil texture and structure affect water retention, aeration, and microbial activity, crucial for nutrient cycling. Understanding soil types guides tailored nutrient management strategies for optimal plant growth.

1.6 Objectives

The objective is to identify optimal feature sets from diverse soil types within a dataset of soil images, leveraging both texture and color features. Following feature extraction, these features will be applied in machine learning models for soil classification, aiming to accurately assign soil samples to respective classes using various machine learning techniques. This classification process contributes to informed decision-making in nutrient analysis, particularly through the utilization of deep learning techniques, thereby enhancing understanding of soil characteristics and facilitating effective agricultural management practices.

CHAPTER 2

METHODOLOGY

2.1 Methodology

To extract the best features and classify the soil images, a computer-aided system is proposed. The proposed methodology consists of image processing techniques such as:

a) Data Acquisition

b) Feature Extraction

c) Machine Learning based Classification

The proposed methodology has been demonstrated in Fig 2 below.

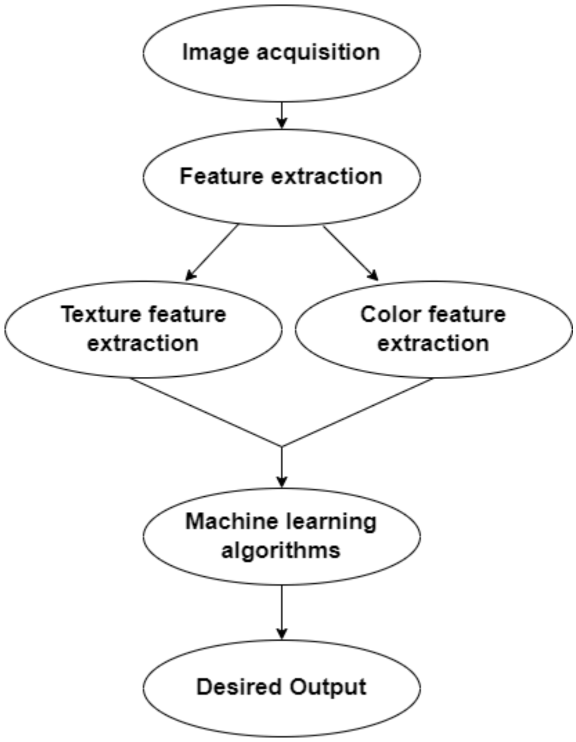

Figure 2: Block diagram of proposed method.

2.2 Image Acquisition

The project began by sourcing soil image datasets from reputable online platform, encompassing five soil types: black, cinder, laterite, peat, and yellow soils, with 50 images per type. The collected images varied in size and format, necessitating preprocessing. Images were resized to 150x150 pixels and converted to JPG format for uniformity. This standardized dataset facilitated compatibility with machine learning algorithms.

2.3 Features extraction

In the feature extraction phase, we utilized a variety of common methods to extract two main types of features from our soil image dataset: texture features and color features. These features are crucial for distinguishing between different types of soil.Both texture and color features are essential for accurately classifying soil types[3].Texture features provide information about the physical properties of the soil surface, while color features offer insights into its visual appearance and chemical composition. By combining these features, our classification model can better differentiate between different types of soil, leading to more accurate classification results.

2.3.1 Texture Features extraction

Texture features were extracted using diverse methods: GLCM, capturing spatial relationships of pixel intensities; Wavelet Transform, decomposing images into frequency components; and Mask Convolution, highlighting specific textural patterns. These methods provided insights into soil surface characteristics such as roughness and smoothness. Texture features are crucial in soil classification as they encode important surface properties, aiding in distinguishing between different soil types based on their unique textural patterns. By incorporating texture features, our classification model gains the ability to accurately identify and classify soil samples, contributing to advancements in agriculture, environmental monitoring, and land management practices.

2.3.1.1 GLCM Texture Feature Extraction :

The Gray-Level Co-occurrence Matrix (GLCM) method was employed to analyze the spatial relationships between pixel intensities in the images. By quantifying the frequency of occurrence of pixel pairs with specific intensity values and distances, GLCM-based features were extracted to characterize textural properties such as roughness, coarseness, and homogeneity.In our project, GLCM yielded 15 features per image: Contrast, Correlation, Energy, Homogeneity, Angular Second Moment, Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation 1, and Information Measure of Correlation 2. Each type of soil had 50 images, resulting in 50 X 15 GLCM features per soil type. These features serve as discriminative metrics, enabling accurate classification of soil samples based on their distinct textural characteristics.

2.3.1.2 Wavelet Transform Texture Feature Extraction :

Wavelet transform techniques were applied to decompose the soil images into different frequency bands or scales. This multi-resolution analysis enabled the extraction of texture features at varying levels of detail, capturing both fine and coarse textural patterns present in the images[4].In wavelet transform feature extraction, 15 features were extracted per image, including mean, standard deviation, entropy, skewness, kurtosis, range, median, median absolute deviation, interquartile range, root mean square, variance, contrast, correlation, energy, and homogeneity. Specifically, for the four coefficients LL, HL, LH, and HH, 16 features (contrast, correlation, energy, homogeneity) were computed for each, totaling 64 features. Additionally, the remaining 11 features were calculated for each coefficient, resulting in 44 features. Overall, 108 features per image were derived, facilitating detailed characterization and discrimination of soil textures during classification.

2.3.1.2 Mask Convolution Texture Feature Extraction :

Mask convolution feature extraction involves convolving soil images with predefined filter masks to highlight specific textural patterns. This method enhances the visibility of distinct features like edges, corners, or blobs within the images. In our project, mask convolution facilitated the extraction of discriminative texture descriptors, aiding in the characterization of soil samples. By accentuating unique textural elements, this technique contributes to the accurate classification of soil types based on their surface patterns[5]. Mask convolution adds valuable information to the feature set, enhancing the robustness and effectiveness of the classification model in distinguishing between different soil textures and structures.

2.3.2 Color Features extraction

Color feature extraction involved capturing various aspects of color distribution in soil images. RGB color features represented color intensity levels in red, green, and blue channels, while mean HSV features captured dominant hues, saturation levels, and brightness variations. CMYK features provided insights into color compositions, and color moment features quantified statistical properties of color distributions. These features are crucial in soil classification as they encode essential visual characteristics, aiding in distinguishing between different soil types based on their unique color signatures. By incorporating color features, our classification model gains the ability to accurately identify and classify soil samples, contributing to advancements in agriculture and environmental monitoring.

2.3.2.1 RGB Color Feature Extraction :

The Red-Green-Blue (RGB) color model was utilized to represent each pixel in the images as a combination of red, green, and blue color channels. RGB-based features were extracted to capture the overall color composition and intensity variations within the soil samples. For RGB color feature extraction, three-channel features are obtained for each image, resulting in 150 color features for a single soil type dataset comprising 50 images. These features encapsulate the intensity levels of red, green, and blue colors present in the soil images. By analyzing the distribution of these color channels, our classification model gains valuable insights into the visual characteristics of different soil types. The extraction of 150 RGB color features per soil type dataset contributes to the comprehensive characterization of soil color variations, enabling accurate classification based on color properties.

2.3.2.2 Mean HSV Color Feature Extraction:

The Hue-Saturation-Value (HSV) color model was employed to represent colors based on their hue, saturation, and value components. Mean HSV values were calculated across the images to characterize the dominant color hues, saturation levels, and brightness variations present in the soil samples[6]. For Mean HSV color feature extraction, three features are derived from each image, totaling 150 color features for a soil type dataset consisting of 50 images. These features represent the mean values of hue, saturation, and brightness components across the images. In classification, Mean HSV features play a crucial role as they encode essential information about the color distribution and composition of soil images. By incorporating these features, our classification model gains the ability to effectively discriminate between different soil types based on their distinct color characteristics.

2.3.2.3 CMYK Color Feature Extraction:

The Cyan-Magenta-Yellow-Black (CMYK) color model, commonly used in printing and color reproduction, was leveraged to represent colors based on subtractive color mixing principles. CMYK-based features were extracted to capture the unique color compositions and tonal variations within the soil images.For CMYK color feature extraction, four features are obtained from each image, resulting in a total of 200 color features for a soil type dataset containing 50 images. These features are derived from the Cyan, Magenta, Yellow, and Black color components, providing insights into the subtractive color mixing process and the composition of colors within the soil samples. CMYK color features are significant in classification as they capture nuanced variations in color composition and tonal properties, enabling the model to distinguish between different soil types based on their unique color signatures. Incorporating CMYK features enhances the classification accuracy by considering additional aspects of color representation.

2.3.2.4 Color Moment Feature Extraction:

Color moment feature extraction involves computing statistical properties such as mean, variance, skewness, and kurtosis of color distributions in soil images. These moments provide insights into the spatial and spectral characteristics of soil color variations. In our project, color moment features encapsulated important statistical information about the color distribution, facilitating the characterization of soil samples based on their color profiles[6].By quantifying the central tendency, dispersion, and shape of color distributions, color moment features contribute to the accurate classification of soil types by capturing subtle nuances in color composition and texture, thus enhancing the discriminative power of the classification model.

2.4 Machine learning algorithms used for soil classification

Machine learning plays a pivotal role in soil classification by automating the process of identifying and categorizing soil types based on their unique features. By leveraging algorithms and statistical techniques, machine learning models can learn from labeled soil image datasets, discerning patterns and relationships within the data to make accurate predictions. This automation not only speeds up the classification process but also enhances its accuracy by minimizing human error and subjectivity. Moreover, machine learning enables the handling of large and complex datasets, accommodating the diverse range of soil characteristics present in real-world scenarios

In this project, several machine learning algorithms were employed to classify soil types and evaluate classification accuracy. These algorithms include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest, Neural Network (NN) classifier, and Kernel classifier. Each algorithm offers distinct advantages in handling different types of data and capturing underlying patterns within the soil image features. By comparing the performance of these algorithms, we aimed to identify the most effective approach for accurately classifying soil types based on their visual and textural characteristics.

.2.4.1 Kth Nearest Neighbor Algorithm

K-Nearest Neighbors (KNN) is a simple yet powerful machine learning algorithm used for classification tasks. It operates based on the principle of similarity, where a new data point is classified by a majority vote of its neighbors. KNN is particularly effective in soil classification as it doesn't assume any underlying distribution of the data and can handle complex and nonlinear relationships between features and classes[7]. In this project, KNN was implemented in MATLAB to classify soil types based on extracted features. The algorithm utilizes the feature vectors obtained from texture and color feature extraction processes. By computing the distances between feature vectors, KNN identifies the k-nearest neighbors of each sample and assigns it the majority class label among its neighbors. The accuracy of soil classification using KNN was evaluated across different feature sets, providing insights into the effectiveness of various feature extraction techniques in discriminating between soil types.

2.4.2 Support Vector Machine Algorithm

Support Vector Machine (SVM) is a versatile supervised learning algorithm widely used for classification tasks. SVM aims to find the optimal hyperplane that separates different classes in the feature space while maximizing the margin between classes[8]. It is effective in handling high-dimensional data and can capture complex decision boundaries, making it suitable for soil classification tasks where features are extracted from soil images.In our project, SVM was implemented in MATLAB to classify soil types based on extracted features. SVM learns to classify soil samples by mapping them into a high-dimensional feature space and finding the hyperplane that best separates different soil types. By adjusting parameters such as the kernel function and regularization parameter, SVM optimizes the classification accuracy. The performance of SVM in soil classification was evaluated using various feature sets, providing insights into its capability to accurately classify soil types based on their extracted features.

2.4.3 Random Forest Algorithm

The Random Forest algorithm is a powerful ensemble learning technique widely utilized for classification tasks. It operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes of the individual trees. Random Forest is robust to overfitting, handles high-dimensional data well, and is resilient to noise and outliers[9]. These characteristics make it well-suited for soil classification tasks where the goal is to accurately classify soil types based on diverse and complex features extracted from soil images. In our project, the Random Forest algorithm was implemented in MATLAB to classify soil types using extracted features. Random Forest builds an ensemble of decision trees, each trained on a random subset of features and data samples. During classification, the algorithm aggregates the predictions from individual trees to make the final decision. By adjusting parameters such as the number of trees in the forest and the maximum depth of each tree, Random Forest optimizes classification accuracy. The performance of Random Forest in soil classification was evaluated across different feature sets, providing insights into its effectiveness in accurately classifying soil types based on their extracted features.

2.4.4 Neural Network (NN) Classifier Algorithm

The Neural Network (NN) classifier is a fundamental machine learning model inspired by the structure and function of biological neural networks. It consists of interconnected nodes organized into layers, including an input layer, one or more hidden layers, and an output layer. NNs are adept at capturing complex nonlinear relationships in data and can learn intricate patterns from large datasets. This makes them well-suited for soil classification tasks where the relationships between extracted features and soil types may be intricate and nonlinear. In our project, the NN classifier was implemented in MATLAB to classify soil types based on extracted features. The NN model learns to classify soil samples by adjusting the weights and biases of connections between nodes through a process called training. During training, the NN iteratively adjusts its parameters to minimize the difference between predicted and actual class labels. By fine-tuning parameters such as the number of hidden layers, neurons per layer, and activation functions, the NN classifier optimizes classification accuracy. The performance of the NN classifier in soil classification was evaluated across different feature sets, providing insights into its capability to accurately classify soil types based on their extracted features.

2.4.5 Kernel classifier Algorithm

The Kernel classifier is a powerful machine learning algorithm used for classification tasks, particularly in scenarios where the data is not linearly separable in its original feature space. It operates by transforming the input features into a higher-dimensional space where the data becomes separable, allowing for the construction of a linear decision boundary. The kernel function plays a crucial role in this transformation, enabling the classifier to capture complex relationships between features. Kernel classifiers are well-suited for soil classification tasks where the extracted features may exhibit nonlinear relationships with soil types.In our project, the Kernel classifier was implemented in MATLAB to classify soil types based on extracted features. The classifier utilizes kernel functions such as polynomial, radial basis function (RBF), or sigmoid to map the feature vectors into a higher-dimensional space. By adjusting parameters such as the kernel type and regularization parameter, the Kernel classifier optimizes classification accuracy. The performance of the Kernel classifier in soil classification was evaluated across different feature sets, providing insights into its effectiveness in accurately classifying soil types based on their extracted features.

CHAPTER 3

RESULT ANALYSIS

3.1 Result obtained from the machine learning algorithms

After implementing the various machine learning algorithms and evaluating their performance, we obtained results in terms of classification accuracy. The accuracy metric provides insights into the effectiveness of each algorithm in accurately classifying soil types based on the extracted features. By comparing the accuracies achieved by different algorithms, we can identify the most suitable approach for soil classification in our project. Additionally, analyzing the impact of different feature sets on classification accuracy allows us to determine which features contribute most significantly to the classification task. These accuracy results serve as valuable feedback for refining the classification model and optimizing its performance for real-world soil classification applications.

3.1.1 Kth Nearest Neighbor Algorithm Accuracy Result :-

Here are the results obtained from the KNN algorithm in terms of accuracy percentage. The table below illustrates the accuracy for various combinations of extracted features:

Table 3.1: Kth Nearest Neighbor Accuracy Result Table

KNN	RGB	HSV	CMYK	COLOR MOMENT
GLCM	70.4%	68.8%	59.2%	59%
WAVELET	71.2%	71.2%	77.6%	59.6%
MASK CON.	66%	62.5%	52.8%	54.3%

3.1.2 Support Vector Machine Algorithm Accuracy Result :-

Here are the results obtained from the SVM algorithm in terms of accuracy percentage. The table below illustrates the accuracy for various combinations of extracted features:

Table 3.2: Support Vector Machine Accuracy Result Table

SVM	RGB	HSV	CMYK	COLOR MOMENT
GLCM	72.8%	69.6%	59.2%	61.0%
WAVELET	72.8%	72.4%	78.0%	64.8%
MASK CON.	64%	60.5%	51.8%	52.3%

3.1.3 Random Forest Algorithm Accuracy Result :-

Here are the results obtained from the Random Forest algorithm in terms of accuracy percentage. The table below illustrates the accuracy for various combinations of extracted features:

Table 3.3: Random Forest Algorithm Accuracy Result Table

RANDOM FOREST	RGB	HSV	CMYK	COLOR MOMENT
GLCM	65.6%	73.6%	56.4%	58.0%
WAVELET	67.2%	73.2%	74.4%	61.6%
MASK CON.	60%	58.7%	59.2%	60.2%

3.1.4 Neural Network (NN) classifier Algorithm Accuracy Result :-

Here are the results obtained from the Neural Network (NN) algorithm in terms of accuracy percentage. The table below illustrates the accuracy for various combinations of extracted features:

Table 3.4: Neural Network Classifier Algorithm Accuracy Result Table

NN CLASSIFIER	RGB	HSV	CMYK	COLOR MOMENT
GLCM	72.8%	66.0%	56.8%	59.0%
WAVELET	74.4%	73.2%	73.6%	61.6%
MASK CON.	57%	59.2%	56.4%	53.3%

3.1.5 Kernel classifier Algorithm Accuracy Result:-

Here are the results obtained from the kernel classifier algorithm in terms of accuracy percentage. The table below illustrates the accuracy for various combinations of extracted features:

Table 3.5: Kernel Classifier Algorithm Accuracy Result Table

KERNEL CLASSIFIER	RGB	HSV	CMYK	COLOR MOMENT
GLCM	65.6%	73.6%	56.4%	58.0%
WAVELET	67.2%	73.2%	74.4%	61.6%
MASK CON.	60%	58.7%	59.2%	60.2%

3.2 Comparative Analysis

Table 3.6: Comparative analysis of ML algorithms

	KNN	SVM	RANDOM FOREST	NN CLASSIFIER	KERNEL CLASSIFIER
RGB & GLCM	70.4%	72.8%	65.6%	72.8%	**80.4%**
RGB & WAVELET	71.2%	72.8%	67.2%	74.4%	**78%**
RGB & MASK CON	**66%**	64%	60%	57%	61%
HSV & GLCM	68.8%	69.6%	**73.6%**	66%	68.4%
HSV & WAVELET	71.2%	**74.1%**	73.2%	73.8%	74%
HSV & MASK CON	**62.5%**	60.5%	58.7%	59.2%	56%
CMYK & GLCM	59.2%	59.2%	56.4%	56.8%	**61.6%**
CMYK & WAVELET	77.6%	**78%**	74.4%	73.6%	77.6%
CMYK & MASK CON	52.8%	51.8%	**59.2%**	56.4%	57%
CLR MO. & GLCM	59%	61%	58%	**62%**	61%
CLR MO. & WAVELET	59.6%	**64.8%**	61.6%	61.6%	63.6%
CLR MO. & MASK CON	54.3%	52.3%	60.2%	53.3%	**61.2%**

In the analysis of various classification algorithms applied to different feature combinations, several notable trends emerge. For RGB and GLCM features, the kernel classifier stands out as the most accurate, achieving an impressive 80.4% accuracy rate, surpassing KNN, SVM, Random Forest, and NN classifiers. Similarly, when considering RGB combined with Wavelet features, the kernel classifier maintains its dominance, achieving the highest accuracy among the five algorithms at 78%. However, for RGB combined with Mask Convolution features, the KNN classifier outperforms the others, attaining a notable 66% accuracy rate.

Moving to HSV features, different combinations yield different top-performing algorithms. In the case of HSV combined with GLCM features, Random Forest emerges as the most accurate classifier, achieving a commendable accuracy rate of 73.6%. Conversely, when HSV is combined with Wavelet features, SVM takes the lead with a 74.1% accuracy rate. Lastly, for HSV combined with Mask Convolution features, KNN exhibits the highest accuracy among the algorithms at 62.5%.

Shifting focus to Color Moment features, distinct algorithmic preferences emerge. When paired with GLCM features, the NN classifier shines with a 62% accuracy rate, showcasing its effectiveness. On the other hand, combining Color Moment with Wavelet features favors SVM, which achieves a notable accuracy rate of 64.8%. Lastly, when Color Moment is combined with Mask Convolution features, Kernel classifier emerges as the top performer, boasting a respectable accuracy rate of 61.2%.

Overall, these findings underscore the importance of feature selection and algorithm choice in achieving optimal classification accuracy, with different combinations yielding varying results across different classifiers.

CHAPTER 4

CONCLUSION AND FUTURE SCOPE

4.1 Conclusion and Future Scope

Looking ahead, our project lays the groundwork for exciting advancements in soil analysis and agricultural practices. Moving forward, we envision delving deeper into soil analysis using advanced deep learning methodologies. This entails expanding our dataset by collecting soil images from diverse locations, enriching our understanding of soil variability and patterns.

With a larger and more diverse dataset, we aim to employ sophisticated deep learning models to analyze soil nutrient content with greater accuracy and precision. By leveraging the power of deep learning algorithms, we can uncover intricate relationships between soil features and nutrient levels, offering invaluable insights for optimizing agricultural productivity and sustainability.

Furthermore, we foresee the development of a user-friendly web application that harnesses the capabilities of our trained models. This application will empower farmers and agricultural professionals to upload soil images effortlessly and receive instant classifications and predictions regarding nutrient levels. By democratizing access to advanced soil analysis tools, we aim to revolutionize soil management practices, fostering informed decision-making and sustainable agricultural practices for the future.

CHAPTER 5

REFERENCES

5.1 References

[1] Hakeem, Khalid & Sabir, Muhammad & Akhtar, Javaid. (2017). Soil Science: Concepts and Applications.

[2] Abdulkadir, Mohamed. (2017). Chapter One-Introduction to Soil.

[3] Pethkar S (2018) Classification of soil image using feature extraction. Int J Res Appl Sci Eng Technol 6:819–823

[4] Zhang X, Younan NH, King RL (2003) Soil texture classification using wavelet transform and Maximum Likelihood Approach. 7929–7931

[5] Shenbagavalli R, Ramar K (2011) Classification of soil textures based on Laws features extracted from preprocessing images on sequential and random windows. Bonfring Int J Adv Image Process 1:15–18

[6] Srivastava, Pallavi & Shukla, Aasheesh & Bansal, Atul. (2021). A comprehensive review on soil classification using deep learning and computer vision techniques. Multimedia Tools and Applications. 80. 10.1007/s11042-021-10544-5.

[7] Maniyath SR, Hebbar R, Akshatha KN, Architha LS, Rama Subramoniam S (2018) Soil color detection using Knn classifier. Proc. - 2018 Int. Conf. Des. Innov. 3Cs Comput. Commun. Control. ICDI3C 2018:52–55

[8] Barman U, Choudhury RD (2020) Soil texture classification using multi-class support vector machine. Inf Process Agric 7(2):318–332

[9] Shukla G, Garg RD, Srivastava HS, Garg PK (2018) An effective implementation and assessment of a random forest classifier as a soil spatial predictive model. Int J Remote Sens 39:2637–2669