Machine Learning Models for Soil Nutrient Analysis

MINOR PROJECT

By

Harsh Parashar (B21EC033)

Koninika Tarafdar (B21EC034)

Under the guidance of

Dr. Salam Shuleenda Devi

Department Electronics and Communication Engineering,

National Institute of Technology, Meghalaya

15th March, 2024

CONTENTS

0.1 INTRODUCTION TO THE PROJECT

0.1.1 MACHINE LEARNING MODELS FOR SOIL NUTRIENT ANALYSIS

Machine learning models have revolutionized various fields, and their application in soil science, particularly for nutrient analysis, holds significant promise. Soil nutrient analysis is crucial for optimizing agricultural productivity, managing environmental sustainability, and ensuring food security. Traditional soil testing methods can be labor-intensive, time-consuming, and costly. However, machine learning models offer efficient and accurate alternatives by leveraging data-driven algorithms to analyze complex soil properties and predict nutrient levels. This introduction explores the potential of machine learning models in soil nutrient analysis, highlighting their benefits, challenges, and implications for sustainable land management and agricultural practices.

0.2 WHAT IS 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 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

0.3 SOIL FORMATION

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. 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.

0.4 TYPES OF SOIL

There are several classifications of soil based on different criteria such as texture, composition, and origin. Here are some common types of soil:

SANDY SOIL

CLAY SOIL

SILT SOIL

LOAMY SOIL

PEATY SOIL

CHALKY SOIL

ALLUVIAL SOIL

BLACK SOIL

LATERITE SOIL

SANDY SOIL

0.5 SANDY SOIL

Figure 1: sandy soil

Sandy soil is characterized by its coarse texture and high drainage capacity due to its large particles. It's often light in color and feels gritty to the touch. While it allows water to infiltrate quickly, it doesn't retain moisture well, making it prone to drought. Sandy soil tends to warm up faster in the spring, facilitating early planting. However, its low fertility necessitates regular fertilization for optimal plant growth. Despite its challenges, sandy soil can support certain crops like carrots, potatoes, and peppers, provided they receive adequate irrigation and nutrients. Proper management strategies such as mulching can help mitigate its drawbacks

Sandy soil typically exhibits low nutrient retention due to its coarse structure, leading to leaching of minerals like nitrogen, phosphorus, and potassium. Its low organic matter content contributes to limited nutrient availability for plant uptake. Consequently, sandy soils often require frequent fertilization to sustain crop growth. However, they do tend to contain higher levels of oxygen, aiding root respiration. Amendments such as compost or organic matter can improve nutrient retention and soil structure over time. Balancing fertilization practices with soil testing is crucial to ensure adequate nutrient levels for healthy plant development in sandy soils.

0.6 CLAY SOIL

Figure 2: clay soil

Clay soil is characterized by its fine particles, giving it a smooth, sticky texture when wet and hard, dense consistency when dry. It tends to retain water exceptionally well due to its high plasticity, which can lead to poor drainage and waterlogging. Its compact nature can also inhibit root penetration and restrict airflow, posing challenges for plant growth. Despite these drawbacks, clay soil offers advantages such as good nutrient retention and high fertility. Its ability to hold onto nutrients can support robust plant growth once adequately managed through techniques like soil amendment and proper drainage.

In terms of nutrient composition, clay soil typically holds onto essential minerals like nitrogen, phosphorus, and potassium more effectively than other soil types. This high nutrient retention capacity stems from its small particle size and strong electrochemical properties, which bind nutrients to soil particles. However, clay soils may suffer from nutrient imbalances and compaction issues, affecting root development and overall plant health. Amendments like organic matter and gypsum can help improve soil structure and nutrient availability, ensuring optimal conditions for plant growth in clay soils. Balancing fertilization practices and soil management techniques is essential to maximize productivity while minimizing nutrient leaching and runoff.

0.7 SLIT SOIL

Figure 3: slit soil

Silt soil falls between sandy and clay soils in terms of particle size, featuring medium-sized particles that offer a balance of drainage and water retention. Its texture is smooth and silky when dry, becoming slick and moldable when moistened. Silt soil offers good fertility and nutrient retention, making it conducive to supporting a variety of plant life. Its fine particles allow for adequate aeration and root penetration, promoting healthy plant growth. However, silt soil can become compacted over time, hindering drainage and root development. Regular soil testing and management practices such as mulching and proper irrigation help maintain its optimal structure for plant growth.

silt soil typically retains nutrients relatively well due to its moderate particle size and surface area. It can hold onto essential minerals like nitrogen, phosphorus, and potassium, providing a fertile environment for plant roots to access nutrients. However, silt soils may require occasional fertilization to replenish nutrients depleted by plant uptake and leaching. Amendments such as organic matter and compost can further enhance nutrient retention and soil structure, promoting healthier plant growth and improved yields. Proper soil management practices tailored to silt soil characteristics ensure optimal nutrient availability and support sustainable agricultural practices.

0.8 LOAMY SOIL

Figure 4: loamy soil

Loamy soil is often regarded as the ideal soil type for gardening and agriculture due to its balanced mixture of sand, silt, and clay particles. It boasts a crumbly texture that feels soft and rich in the hand, providing excellent drainage while retaining moisture effectively. Loamy soil supports robust root development and facilitates nutrient uptake for plants, making it highly fertile and conducive to various crops. Its well-aerated structure allows for optimal oxygen exchange, promoting healthy microbial activity and decomposition of organic matter. Gardeners and farmers favor loamy soil for its versatility and ability to sustain plant growth across diverse climates and conditions.

In terms of nutrient composition, loamy soil is renowned for its excellent nutrient retention capabilities. Its balanced combination of sand, silt, and clay particles creates an optimal environment for holding onto essential minerals like nitrogen, phosphorus, and potassium. This nutrient-rich soil provides a steady supply of nutrients to plants, reducing the need for frequent fertilization. Additionally, loamy soil tends to have a high organic matter content, further enhancing its fertility and soil structure. With proper soil management practices, including crop rotation and organic amendments, loamy soil can sustain productive yields while promoting long-term soil health and sustainability

0.9 PEATY SOIL

Figure 5: peaty soil

Peaty soil is characterized by its high organic matter content, primarily composed of decomposed plant material accumulated over thousands of years in waterlogged environments. It has a dark brown to black color and a spongy texture, retaining moisture exceptionally well. Peaty 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, peaty soil is prized for its fertility and ability to support unique plant communities, including acid-loving species like rhododendrons and blueberries.

peaty 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.

0.10 CHALKY SOIL

Figure 6: chalky soil

Chalky soil, also known as limestone soil, is characterized by its alkaline pH and high calcium carbonate content, derived from the weathering of limestone bedrock. It typically has a light, crumbly texture and tends to drain well, making it prone to drought. Chalky soil often appears white or pale due to its calcium-rich composition. While it offers good drainage, it may suffer from nutrient deficiencies, particularly in acidic-loving plants. Despite its challenges, chalky soil supports unique plant species adapted to alkaline conditions, such as lavender and thyme, and is well-suited for growing certain crops like grapes and barley.

chalky soil tends to have limited nutrient availability due to its alkaline nature and high calcium content. It may lack essential minerals like iron, zinc, and manganese, which can lead to nutrient deficiencies in plants. Gardeners and farmers often amend chalky soil with organic matter or acidic fertilizers to improve nutrient retention and adjust pH levels for optimal plant growth. While managing chalky soil requires careful consideration of plant selection and soil amendments, it can be productive when properly cultivated and balanced with appropriate fertilization practices.

0.11 BLACK SOIL

Figure 7: 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. 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.

0.12 ALLUVIAL SOIL

Figure 8: Alluvial soil

Alluvial soil is one of the most fertile and agriculturally productive soil types, formed by the deposition of sediment carried by rivers and streams. It is commonly found in river valleys, floodplains, and deltas, where periodic flooding deposits nutrient-rich sediments. Alluvial soil exhibits excellent drainage properties and a loamy texture, making it suitable for a wide range of crops. Its fertility is attributed to the accumulation of organic matter, clay, silt, and minerals transported by water over time. Alluvial soil supports intensive agriculture and is widely used for growing staple crops such as rice, wheat, and maize. However, its susceptibility to erosion and compaction necessitates sustainable land management practices to maintain soil health and productivity.

Alluvial soil is rich in nutrients essential for plant growth and productivity. Its fertility is derived from the deposition of organic matter, clay, and minerals carried by rivers and streams. Alluvial soil typically contains high levels of nitrogen, phosphorus, and potassium, crucial for supporting vigorous plant growth and crop yields. Additionally, micronutrients such as calcium, magnesium, and sulfur are abundant in alluvial soil, contributing to overall soil fertility. The continuous deposition of sediment during floods replenishes nutrients in alluvial soil, ensuring its productivity over time. However, intensive agriculture and improper land management practices can lead to soil degradation and nutrient loss. Implementing soil conservation measures such as contour farming, cover cropping, and erosion control is essential for preserving the fertility and sustainability of alluvial soil for future generations

0.13 LATERITE SOIL

Figure 9: laterite soil

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.

0.14 LITERATURE REVIEW

Here are the machine learning models used for soil nutrient analysis in last 5 years

0.14.1 MACHINE LEARNING MODELS FOR SOIL NUTRIENT ANALYSIS

0.15 2024

Using a transformation method on the initial soil picture. Applying statistical measures on both the source photograph and the altered image, characteristics of texture, colour, and shape are retrieved. Following the calculation of the Euclidean method's distance formula, classification is fulfilled.Concluding with the ultimate step of calculating the total number of accurate checks retrieved from the database.

Acquisition: We acquired the images required for the testing and training datasets from the web. The standard pixel size that has been utilized here is 256 x 256, has been kept uniform through

Detection: Colour is a property that affirms details about a place's age, mineral composition, presence of humidified organic materials, supported crops, and other characteristics Utilizing PCA analysis, pictures are classified during the testing stage, and index values and training variables are contrasted. The final pH is determined

Data processing: another feature introduced another classification method that uses multi-SVM and a linear kernel function

0.15.1 MULTI-SVM

Multi-class SVM (Support Vector Machine) is a machine learning method used in soil classification, where soil types are categorized into multiple classes based on various features like texture, color, and composition. SVM aims to find the optimal hyperplane that maximizes the margin between classes while minimizing classification errors. It works by transforming data into a high-dimensional space where classes are separable. Through iterative optimization, SVM learns the decision boundaries between different soil classes, allowing accurate classification of new soil samples. This method is robust, efficient, and widely used in soil science for its ability to handle complex classification tasks with high-dimensional feature spaces.

0.15.2 LINEAR KERNEL FUNCTION

In soil classification, the linear kernel function is a fundamental component of Support Vector Machines (SVM). It's used to map input features to a higher-dimensional space where classes are linearly separable. The linear kernel calculates the dot product between feature vectors, effectively measuring the similarity between samples. This kernel is particularly suitable when soil classification involves linearly separable classes or when the dataset is high-dimensional. By employing the linear kernel, SVM can efficiently find the optimal hyperplane that maximizes the margin between different soil types, leading to accurate classification results without the need for complex transformations or computations.

CNN architecture: In this project, we have applied 32 layers wherein 2 convolutional layers are present. The model's foundational layer is a 2D convolutional layer (Conv2D) with a (3, 3) size kernel, and identical padding. The input form of the pictures in this layer is (256, 256, 3), which is the typical size of a jpg image, further described better in Fig.3. It also features a rectified linear unit (ReLU) activation function. A maxpooling layer (MaxPooling2D) with a pool size (2, 2) is drawn upon. Ultimately, the size of dense layers processed for output is 5.

0.15.3 LIGHT-SOILNET NETWORK ARCHITECTURE

Image pre-processing and augmentation: The dimensions of the soil images captured are 2992x2992, and the size of the images is 2.25 MB on average. To reduce the computational complexity of the proposed model, an optimal image size 728x728 has been chosen where the loss of information is less when compared with the image of size 512x512. In this stage, the region of interest (ROI) of the images has been extracted to distinguish the soils from the background. The HSV color threshold technique has been applied to extract the ROI of an image shows the acquired and ROI-extracted images of the soil. The tool used to extract ROI from images is MATLAB2021a.

The proposed Light-SoilNet architecture comprises multiple layers for the classification of soil by reducing the number of parameters and increasing the model's efficiency. The layers used in the architecture are the convolution layer, ReLU, batch normalization, max-pooling, dropout, fully connected, and softmax layer.

The images are converted into greyscale to perform an adaptive histogram because of the significant color shift when performing the histogram equalization independently on the images red, green, and blue components. The adaptive histogram images enhance the edges of the images, which helps identify the texture of the soil images

The convolutional layer : contains filters whose parameters need to be learned to perform the convolutional operations on the input images. 2-D convolutional operation is used for two-dimensional soil images. The smaller size kernel filters are used in the proposed model to extract the spatial features: edges and corners of the soil images

Batch Normalization : To speed up the training of the SoilNet model, batch normalization is used after the convolutional layer. It improves the performance and stability of the training model, reduces the covariate shift and achieves efficient learning in the model.

Pooling : In the Light-SoilNet network, the max-pooling layer of size 2×2 and 3×3 with stride two is used to reduce the network depth and learnable parameters. Max-pooling layer with convolution layers helps in position invariant feature detection.

0.16 2023

0.16.1 DECISION TREE

DT represented as a tree structure, with a leaf node indicating the target attribute and a decision node defining a test to be conducted on a specific characteristic of the given observation, with one branch for each potential test outcome, The DT classifier proceeds from observations (represented as branches) to inferences about the target value (represented in the leaves). Examining the value in the decision node begins the categorization process and continues until a leaf node is reached . The confidence factor determines the level of pruning adversity; higher levels will result in less pruning. The proportion of the data used to prune the tree depends on the number of folds, and the remaining data are used for tree growth. Insufficient folds will result in more overfitting. A seed value that is chosen at random helps in minimizing error pruning. The DT can be generated from the dataset by establishing the attributes for each node and the conditions for splitting at that node. The algorithm selects "best" to minimize the step's information gain. It accomplishes this by employing equations based on entropy and aiming to reduce entropy

0.16.2 K NEAREST NEIGHBOURS

is used for both classification and regression, where the algorithm memorizes the entire training dataset and uses it to make predictions for new data points. For Classification, during training, KNN uses the feature vectors and their corresponding class labels from the training dataset. KNN selects the K nearest data points by calculating the distance between the new data point and all data points in the training set. For classification, it counts the class labels of these K neighbors and assigns the class label that occurs most frequently as the predicted class label for the new data point. For Regression, it calculates the average (mean) of the target values of these K neighbors and assigns the mean as the predicted target value for the new data point

0.16.3 SUPPORT VECTOR MACHINES

SVM can be used for regression and classification tasks. For classification, it utilizes the concept of decision planes to divide data into appropriate categories using decision boundaries. A specific feature vector is built from encoded representations of residue attributes for each sequence. The SVM parameters are calibrated by first translating the input vectors into a high-dimensional space. It seeks to locate the separating hyperplanes that increase the margin between data sets given a collection of data points in an N-dimensional space for each available class. It can use polynomial or radial basis function kernels to map the data into a higher-dimensional space where separation is possible. For regression, it focuses on fitting a hyperplane that captures a specified margin of error around the data points. A set of input–output pairs, where each input has a corresponding target value. It aims to find the hyperplane that minimizes the deviation of predicted values from actual values within the specified margin of error. It can also utilize kernel to handle non-linear relationships between inputs and targets. The SVM can handle linear and nonlinear data and outperforms ANN and DT without a large training data set.

0.16.4 ARTIFICIAL NEURAL NETWORK

are powerful analytical models based on biological neural networks. The ANN model can be constructed by joining many weighted connections called neurons. The neurons are normally organized in a layer or vector that can be connected to neurons in the next layer. A connection weight, wn,m, multiplies the input values to a processing element to emulate the strengthening of neural networks. Learning in ANNs is simulated by adjusting connection weights. The ANN neurons use the dataset's nonlinear functions as input to produce the outputs. It allows us to minimize errors by altering neurons' weights and biases . ANNs are very useful for modeling nonlinear systems. ANNs can perform classification using knowledge gained through repeated experience, similar to human learning .

0.17 2022

0.17.1 MULTI LAYER PERCEPTRON ARTIFICIAL NEURAL NETWORKS

,it is also called feedforward neural networks or deep feedforward networks, is a popular machine learning technique that is based on biologically the process of MLP-ANN is basically structured by neurons. A neuron takes in one or more inputs, multiplies each input by a weight , sums the weighted input's values along with some bias value and then feeds the value into an activation function This output is then sent forward to the other neurons deeper in the neural network (if they exist). MLPANN contains three types of layers of neurons: input layer, hidden layers, and output layer that are connected as a network form

0.17.2 RANDOM FOREST

Random Forest is a powerful machine learning algorithm widely employed in soil classification tasks. It operates by constructing multiple decision trees during training, each tree trained on a random subset of the data and features. In soil classification, Random Forest can efficiently handle complex relationships between soil attributes such as texture, pH, and organic matter content. By aggregating the predictions of multiple trees, Random Forest improves robustness against overfitting and enhances generalization performance. It's particularly effective for handling high-dimensional datasets and can provide valuable insights into the relationships between different soil types and their attributes, making it a popular choice in soil science research and applications.

0.17.3 SUPPORT VECTOR MACHINES

Support Vector Machine (SVM) is a robust algorithm widely used in soil classification due to its effectiveness in handling complex data distributions and high-dimensional feature spaces. In SVM-based soil classification, the algorithm aims to find the optimal hyperplane that separates different soil types while maximizing the margin between classes. It works by transforming input data into a higher-dimensional space where classes are linearly separable, or by using kernel functions to handle non-linear relationships between soil attributes. SVM's ability to handle multi-class classification and its robustness to outliers make it particularly suitable for soil classification tasks where data may be noisy or overlapping. By learning the decision boundaries between different soil types, SVM provides accurate and reliable classification results, making it a valuable tool in soil science research and applications.

0.18 2021

0.18.1 MAXIMUM LIKELIHOOD ESTIMATION

The Maximum Likelihood Estimation (MLE) model was used for soil classification by employing wavelet transform to identify different soil textures.

- Wavelet Transform: This method was applied to extract features from soil images, which is crucial for identifying various soil textures.
- Feature Optimization: During the classification process, Fisher's Linear Discriminant Analysis (FLDA) was used to optimize and reduce the dimensionality of the feature vector.
- Classification: The MLE model then used these optimized features to classify soil textures with high performance, particularly for silt and sand.

The MLE model, in conjunction with wavelet transform and FLDA, provided a systematic approach to classify soil textures effectively. The classification rates for clay, sand, and silt were notably high, demonstrating the model's efficacy in soil texture classification.

0.18.2 EUCLIDEAN DISTANCE

Euclidean distance was used as a metric in the process of soil classification to quantify the similarity between feature vectors representing different soil samples. Specifically, it was employed in the context of Soil Redoximorphic Features (SRFs) quantification and identification system. The system aimed to identify and quantify SRFs from soil core images, and Euclidean distance helped in comparing the color features of these images to a reference, likely the Munsell soil color chart, to classify the soils accurately. The overall accuracy of the system using Euclidean distance for SRFs identification was reported to be 99.61%. This high level of accuracy indicates that Euclidean distance was an effective measure for comparing the color features of soil samples in this study.

0.18.3 SUPPORT VECTOR MACHINE

SVM (Support Vector Machine) was used for soil classification by employing textural features extracted from soil images.

- Feature Extraction: Textural features of the soil images were extracted using methods like gray-level run length matrix1.
- Classifier: SVM, with a linear kernel function, was used as the classifier for the soil classification task.
- Performance: The SVM classifier achieved an accuracy rate of 95.21% for soil classification in the West Guwahati region.

The SVM classifier's ability to handle high-dimensional data and find the optimal hyperplane for classification makes it well-suited for tasks like soil classification based on textural features. The high accuracy rate indicates that SVM was effective in classifying the soil types in the study area.

0.19 2020

0.19.1 ARTIFICIAL NEURAL NETWORK

The study utilized multilayer perceptron artificial neural networks for classifying tilled soil images. Different network topologies were tested to find the best structure for image classification.

- Texture Analysis: Various statistical techniques were employed to analyze the texture of soil images, which is crucial for the classification process. These included gray level co-occurrence matrix, gray level run length matrix, first order statistics of image histogram, and local binary pattern.

- Feature Selection: A data mining procedure by CfsSubsetEval was used for feature selection to reduce the complexity of the input vector for the ANN and improve the performance of the classification process.

The study aimed to optimize the camera height for image acquisition and found that a height of 60 cm provided the best results for the ANN classifier in determining tillage quality. The overall accuracy of the ANN classifier was 72.04% for images taken at this height. The technique is considered feasible for implementation in variable rate secondary tillage machines.

A method for soil classification based on the color of the soil using an Artificial Neural Network (ANN) was utilised.

- Soil Color Analysis: The study utilized soil color values in RGB (Red, Green, Blue) format as input data for the ANN model.

- Neural Network Architecture: A neural network with one hidden layer consisting of 10 neurons was proposed. The input layer had three nodes corresponding to the RGB values, and the output node was designed to predict the soil pH value.

- Data Collection: Secondary data, which included 50 soil samples with known RGB values and pH levels, were used to train the neural network.

- Model Training and Evaluation: The ANN was trained with the collected data, and its performance was evaluated using statistical indicators like Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Determination Coefficient (R^2).

The study aimed to estimate soil pH from the color data, which could be beneficial for agricultural applications such as determining suitable crops for the soil. The results indicated that the ANN model performed well within the specific range of soil colors studied.

0.20 2019

0.20.1 GRADIENT BOOSTING

The study utilized a dataset comprising 805 soil samples obtained from soil drillings during the construction of the new Gayrettepe–Istanbul Airport metro line.

The dataset had both missing data and class imbalance. To handle missing data, KNN imputation techniques were applied. Synthetic minority oversampling technique (SMOTE) was used to balance the dataset.

The study evaluated several ML algorithms using 10-fold cross-validation. Notably, they explored new gradient-boosting methods:

- XGBoost
- LightGBM
- CatBoost

High classification accuracy rates of up to +90% were observed. The accuracy of prediction significantly improved compared to previous research.

In summary, this study demonstrated the effectiveness of ML techniques, especially gradient-boosting methods, for soil classification. These models can enhance the efficiency of soil engineering processes and reduce reliance on costly and time-consuming laboratory tests.