Mathematical Statistician and Engineering Applications
ISSN: 2326-9865
DOI: https://doi.org/10.17762/msea.v70i1.2316

Machine Learning-Based Soil Classification

Kajal Aggarwal

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Article Info
Page Number: 340-347
Publication Issue:
Vol. 70 No. 1 (2021)

Article History
Article Received: 25 January 2021
Revised: 24 February 2021
Accepted: 15 March 2021

Abstract
A crucial component of agriculture is soil. There are several varieties of dirt. Different properties may be found in each kind of soil, and various crops can be grown on various types of soils. To understand which crops do better in different soil types, we need to be aware of their features and traits. In this situation, machine learning approaches may be useful. It has made significant development in recent years. In the realm of agricultural data analysis, machine learning is still a young and difficult study area. In this study, we provide a model that predicts soil series with regard to land type and, in accordance with prediction, suggests appropriate crops. For soil classification, a number of machine learning techniques are utilised, including Decision Tree (CART), Multilayer Perceptron (MLP) and support vector machines (SVM) using a gaussian kernel. The suggested SVM-based technique outperforms several current methods, according to experimental data. The support vector machine technique is used to examine and compare the classification outcomes of various training sample numbers. The purpose of this work is to evaluate the viability of land cover classification using small samples using SVM and, in accordance with the machine learning algorithm, to investigate novel techniques for quick, non-destructive, and precise land cover classification.

1. Introduction

Data mining is the process of finding hidden patterns in massive datasets and connecting them to solve problems via data analysis. Data mining's introduction to the agriculture sector has benefited research. To establish the principles in every scientific discipline, classification is crucial. Finding the differences between the ideas and things may be made easier. Additionally, it offers the data required for doing research in a methodical way. One of the most important elements in an agricultural area for producing crops is the soil. Philosophies for classifying soils are based on empirical data and real-world situations. A connection is made between soil samples and different types of natural entities on the earth's land surfaces thanks to soil categorization.

The constituents of soil include sand, clay, silt, peat, and loam. However, in the target zone, silt, sand, and clay are the most common soil constituents, whereas loam and peat are hardly noticeable. The exploration location has a lot of clay present among the widely distributed soil particles. Clay particles are present in the surface because there is a lot of food being provided. The spaces in the clay type soil allow for the retention of water. The clay particles on the surface hold onto the water for a longer period of time during the winter, preventing drying, and during the summer, it maintains more moisture, causing a supply of nutrients. These soils and their varieties provide a significant difficulty for farmers since the summer causes the surface to break. It is exceedingly difficult to keep these soils fit for farming. Iron and aluminium are

abundant in laterite, which is a form of soil mostly made up of rock type deposits from hot climates. Due to its high iron oxide content, laterite is one of the soil types that is red in hue. Nearly all laterites are rusty-red in hue due to the high iron oxide concentration. The proper creation of the soil surface is ensured by recurrent rainfall and the sunny season. The crops can get enough nutrients from this kind of soil. Farmers have the ability to boost agricultural productivity by studying favourable situations, which will avoid crop damage and loss caused by unfavourable ones. Every day, new plant hybrid kinds are developed from existing ones. On the other hand, as compared to the crop that is cultivated organically, these sorts are deficient in important nutrients. These artificial approaches frequently result in a degradation of the soil quality, which contributes to an even worse deterioration of the ecosystem. The protection of financial losses constitutes the primary focus of the bulk of these artificial solutions. On the other hand, farmers that have a comprehensive awareness of the numerous factors may be able to decrease crop loss and increase their overall production.

According to a poll, between 16 and 20 percent of all the crops produced in India each year are wasted. In India, crops are grown on more than 215.6 million acres (82.6 million hectares) of farmland, 48.92 lakh hectares of which are in Tamil Nadu. Around 46,570.25 acres are utilised for mulberry farming in this area. Tamil Nadu is the second-largest producer of silk cocoons, after Karnataka. One of the most important aspects of agricultural production sustainability is managing soil fertility. This makes it possible to guarantee that the soil has the capacity to provide the plants the necessary nutrients at the appropriate time, in the appropriate amount, and in the accessible form. The amount of micro- and macronutrients in the soil may be determined by their presence or absence. Along with the micronutrients required for the plant enzymatic systems that should be accessible in traces (mg/kg), there are 16 nutrients that are necessary in medium or high concentrations for plant development. Throughout the whole plant life cycle, the micronutrients are essential. When there is a lack of micronutrients in the soil, aberrant plant development may result, resulting in total crop loss. A lot of people are now interested in developing sustainable sericulture since it relies less on outside inputs and can help solve the significant problem of soil erosion. If a certain nutrient is not present in the soil, the signs of a deficit will be obvious. The stability of the worm is significantly influenced by the leaf quality when mulberry is grown for silkworms. The findings are important since mulberry leaves are the only food source for silkworms, as stated.

The primary goal of soil management is to maintain and improve the dynamic soil characteristics in order to increase crop yield. Particularly in emerging nations like India, declining traditional methods of soil management, physical constraints, and population pressure have led to a decline in soil fertility. The fundamental factor ensuring a highly productive system in contemporary agriculture is crop health, which is closely related to soil quality. Adoption of appropriate crop health management techniques may assist in obtaining a steady improvement in crop output. An increase in output may be accomplished by enhancing the quality of the soil through the use of micronutrients as a kind of remediation and by effectively managing the resources contained within the soil. The issues that are linked with crop yield points can potentially be managed by timely detection and control by agricultural

professionals, decision-makers, and farmers for optimal crop environment and soil resource management. This will allow for the problems to be rectified.

Several machine learning (ML) methods and models are now being utilised to solve categorization and prediction problems. The use of ML approaches also significantly helps professionals in the agriculture field handle their issues. The classification and regression methods are effective for predicting agricultural productivity. These models may be used for estimation, yield mapping, supply and demand matching, and general crop management. The machine learning algorithms provide significant advantages for several agricultural and farming applications, including livestock prediction, water management, soil management, crop failure detection, and many more. The collection of real-time data from the soil and crop is done via the usage of a private internet repository. The dataset's values are utilised for the model's training and validation, as well as for evaluation and computation of the outcomes. While building the model, the dataset is divided into training and testing data for this reason. By analysing the soil and using several machine learning algorithms based on the chemical characteristics and micronutrients contained in the soil, the fertility of the soil is rated. ML models are also used to analyse agricultural production and planting. Machine learning methods including linear regression, Naive Bayes, decision trees, artificial neural networks (ANN), and support vector machines (SVM) are used by the current systems. In recent study, machine learning algorithms are employed to assess the soil quality used in clay soil expansion.

2. Literature Survey

T. Islam et al in [1] their proposal for a deep neural network method to crop selection and yield prediction in Bangladesh. The authors train a deep neural network model using information on past crop yields, soil types, and meteorological variables in order to forecast the production of various crops in upcoming growing seasons. The study addresses the model's accuracy and the advantages of using this strategy for crop management in Bangladesh.

An IoT-based real-time energy metering system was designed, according to S. Sheeba Rani et al [2]. The authors create a system that can track and record energy use in real-time using a mix of hardware and software components. The system's technical specifications are covered in depth, along with the system's possible uses in different contexts.

Ajdadi FR et al [3] investigated how machine vision may be used to categorise the size of soil aggregates. The authors create a machine vision system that categorises soil particles according to size using image analysis methods. The system's technical specifications are covered in depth, and field test results are shown to show the system's efficacy.

A neural network technique was suggested by Aziz MM et al [4] to calculate the pH of soil using just its colour. To train a neural network model, the scientists employ a collection of soil samples with known pH values and their related colours. The model's precision and prospective uses for soil analysis are discussed in the publication.

A data mining technique was proposed by Devi M. et al [5] for the study of soil data and agricultural production prediction. The authors create a model for projecting agricultural yields

in future seasons using information on climate, soil type, fertiliser use, and historical crop yields. The model's technical specifications are covered in depth, and experiment results are shown to show the model's efficacy.

A research on the use of data mining tools to analyse agricultural production data was given by Ramesh et al [6], to find valuable patterns and create rules for forecasting agricultural yields, the authors used association rule mining and decision trees. The research demonstrated how data mining methods may help boost agricultural output and decision-making.

Using QuickBird footage, Lu, Hetrick et al [7] carried out land cover categorization research in an intricate urban-rural area. The authors used an object-based method and a maximum likelihood classifier to segment images. According to the research, the object-based technique performed more accurately and effectively than the conventional pixel-based categorization method.

A research on measuring crop development at field sizes using the fusion of Landsat and MODIS data was presented by Gao et al [8]. The scientists created a crop progress index that covers the phases of crop development using spectral and temporal data from the two satellite sensors. The research illustrated the possibility for monitoring and managing crops using remote sensing data.

In research on the best feature combinations for linear discriminant analysis (LDA) classification of functional near-infrared spectroscopy (fNIRS) data in a brain-computer interface application, Noman et al [9] published their findings. The classification of four distinct mental activities was done by the authors using feature selection approaches and LDA. According to the research, selecting the best features increased the fNIRS signals' ability to be classified accurately.

In research published in 2016, Hong et al [10] used functional near-infrared spectroscopy to decode four distinct sound categories in the auditory cortex. The fNIRS signals belonging to various sound categories were classified by the authors using a machine learning method based on support vector machines. The research showed the potential of fNIRS signals for applications in brain-computer interfaces that include auditory processing.

3. Proposed Model

The "Support Vector Machine" (SVM) technique may be used to resolve classification and regression issues. On the other hand, classification issues commonly use this tactic. Each feature's value correlates to a place in n-dimensional space, where n is the number of features present. The SVM method shows all characteristics as coordinates in n-dimensional space. The dataset is classified by locating the hyper-plane that separates it into its several groups according to the nitrogen value. Based on the NPK value of this classification, it is possible to suggest a crop. SVM, one of the most popular binary classifier techniques, is employed in a number of fields, including the classification of NPK. By foreseeing nearby values and categorising them as Low, High, or Medium. Understanding the soil fertility level and crop fertiliser requirements is important for using the model. We can see the classification of the soil

based on village by training the SVM using the training dataset and testing the model's validity with the test dataset. The decision tree is a tool used in multi-step or hierarchical approaches (tree structure). Every branch must start from the base in order to flourish. A root may be compared to an observation. CART is the decision tree algorithm that is used the most often. To get the best possible outcomes, the CART algorithm must first prune the trees. Regression and classification both employ the iterative partitioning method CART. All prediction variables are divided into subgroups via CART, which then creates two successive sub-nodes from the whole data set. The desired variable should fluctuate as little as feasible.

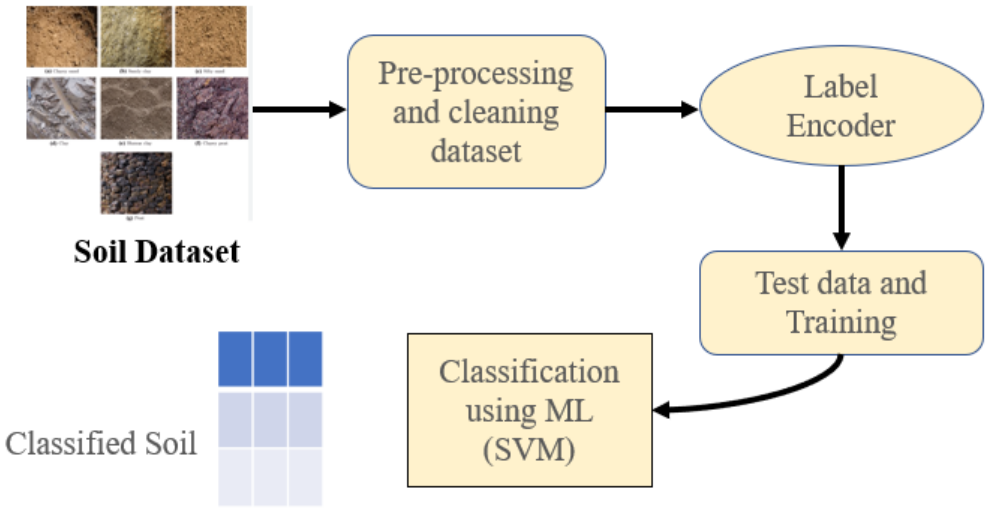


Figure 1. Architecture of proposed Soil classification using ML

categorization using multilayer perceptron's A technique for teaching a function while being watched over is the multiple-layer perceptron. f(.): RxRo is used for training, where x is the number of input dimensions and o is the number of output dimensions. The MLP is composed of several layers of processing units that resemble neurons. The nodes in each layer were connected to those in the one preceding it. Strength and weight of a node might be symmetric or asymmetric. A network has input and output layers, and data moves back and forth between them. Artificial neural networks have employed multilayer perceptron (MLP) as a classification technique. The "Normalizer" node was used to normalise the input data's z-scores before algorithm training. The same process was then applied to real test data. Each layer of nodes in a directed graph connects to the one before it, creating a nonlinear statistical model. There are three different sorts of layers: hidden, input, and output.

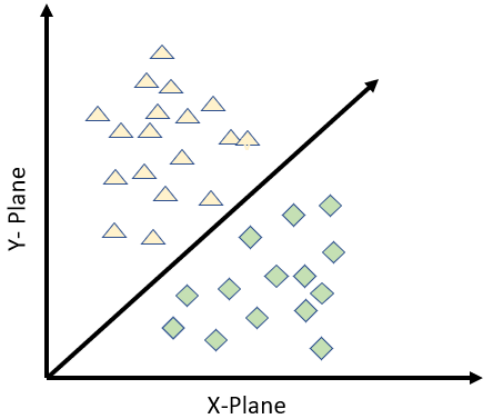


Figure 2. SVM Classification

Humans gain the talent of categorization from reputable sources over the course of many lifetimes, and they then utilise it frequently to carry out their task. Deep learning algorithms in the field of artificial intelligence carry out the same task by picking up knowledge from unstructured input such as texts, sounds, photos, and videos. The distinction between deep learning and machine learning is whether automatic feature extraction or direct learning from the provided data is used. Machine learning is the automatic discovery of important data patterns. It is a branch of artificial intelligence with the goal of employing intelligent software to make it possible for machines to do tasks with skill. Deep learning methods are often developed and put into practise using artificial neural networks. The number of layers in a network determines its strength; the deeper the network, the more layers there may be. Data, model, and algorithm are deep learning's three main building blocks. There may be hundreds of hidden layers in deep learning models. The network's tiers all carry out concurrent non-linear operations. With the help of neurons, connections between the layers are made possible; each layer processes the output of the one before it and sends the results to the next layer for further processing. Network complexity rises as the number of layers rises. Deep neural networks allow the computer to learn entirely from the data without any human input. Giving the network the right and required input is essential to getting the greatest results from it. Although many other types of 3D photos may be used as input, stereo images provide superior results because they have the benefit of being able to extract 3D information from a scene without being impacted by changes in lighting.

Algorithm:

1. Normatively transform soil data from diverse sources.

2. Submit a C or X for each:

Leave-one-out cross validation, version 2.1

2.1.1. SVM training and process testing.

Save and keep a record of the success rate.

Calculating the average success rate (section 2.2).

2.3. If required, update the C and X.

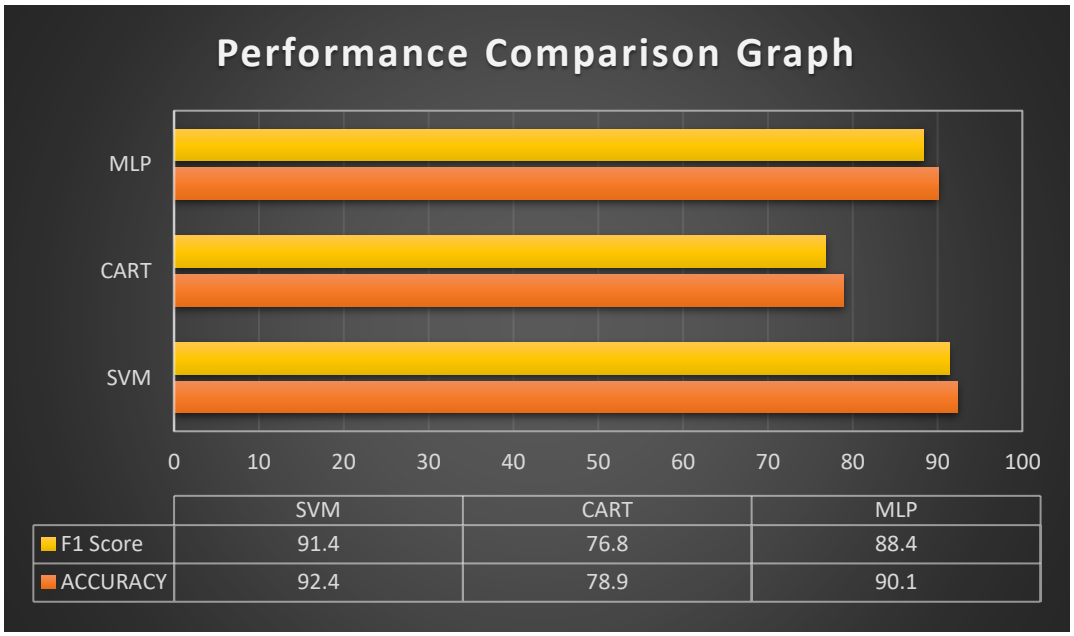
2.4. Return to step 2.1 and add more C and X.

3. To proceed to step 2, choose C or X with an exact and precise average success rate.

using a fine scale within predetermined limits.

4. Result Analysis

Figure 2 shows the accuracy and F1score comparison graph. SVM shows better accuracy compare to other algorithms



	SVM	CART	MLP
F1 Score	91.4	76.8	88.4
ACCURACY	92.4	78.9	90.1

Figure 2. Performance Comparison Graph

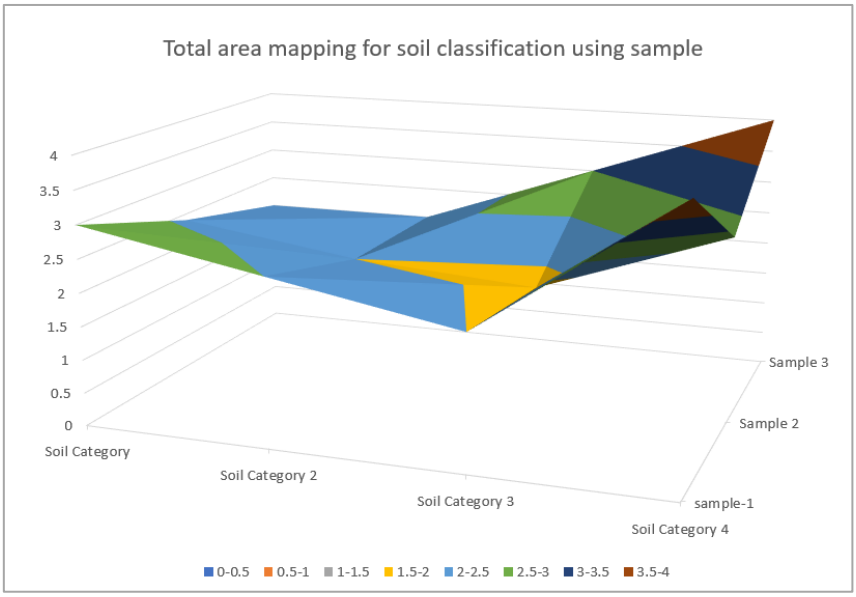


Figure 3 Total area mapping of soil classification using samples

5. Conclusion

A programme for categorising soil types and offering the appropriate crop recommendations for the various soil series. By suggesting the best crop taking into account numerous features, the proposed study will assist farmers in maximising agricultural output, reducing soil degradation in cultivated areas, and reducing fertiliser consumption in crop production. The suggested task assists framers in sustainably choosing the crop for production. The examination of a soil dataset using data mining methods is the goal of this study. It focuses on classifying soil using the many available algorithms. Regression analysis is used to predict untested properties, and automated soil sample categorization is put into practise. SVM machine learning technique enabled superior soil categorization outcomes.

Reference

1. T. Islam, T. A. Chisty, and A. Chakrabarty, "A deep neural network approach for crop selection and yield prediction in Bangladesh," in IEEE region 10 humanitarian technology conference (R10-HTC), pp. 1–6, Malambe, Sri Lanka, 2018.
2. S. Sheeba Rani, K. C. Ramya, V. Gomathy, G. Radhakrishnan, and S. R. B. Prabhu, "Design of IoT based real time energy metering system," International Journal of Innovative Technology and Exploring Engineering, (IJITEE), vol. 8, no. 6S3, 2019.
3. Ajdadi FR, Gilandeh YA, Mollazade K, Hasanzadeh RPR (2016) Application of machine vision for classification of soil aggregate size. Soil Tillage Res 162:8–17
4. Aziz MM, Ahmed DR, Ibrahim BF (2016) Determine the Ph. Of soil by using neural network based on Soil's colour. Int J Advanced Res Comput Sci Software Eng 11:51–54.
5. Devi, M. P. K., Anthiyur, U., & Shenbagavadivu, M. S. (2016). Enhanced Crop Yield Prediction and Soil Data Analysis Using Data Mining. International Journal of Modern Computer Science, 4(6).
6. Ramesh, D., & Vardhan, B. V. (2013). Data mining techniques and applications to agricultural yield data. International journal of advanced research in computer and communication engineering, 2(9), 3477-3480.
7. D. Lu, S. Hetrick, and E. Moran, "Land cover classification in a complex urban-rural landscape with QuickBird imagery," Photogrammetric Engineering & Remote Sensing, vol. 76, no. 10, pp. 1159–1168, 2010.
8. F. Gao, M. C. Anderson, X. Zhang et al., "Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery," Remote Sensing of Environment, vol. 188, pp. 9–25, 2017.
9. N. Noman, F. M. Noori, N. K. Qureshi, and H. Keum-Shik, "Determining optimal feature-combination for LDA classification of functional near-infrared spectroscopy signals in brain-computer interface application," Frontiers in Human Neuroscience, vol. 10, p. 237, 2016.
10. K.-S. Hong and H. Santosa, "Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy," Hearing Research, vol. 333, pp. 157–166, 2016.