SOIL TEXTURE CLASSIFICATION AND SEGMENTATION USING HYBRID RANDOM FOREST WITH ARTIFICIAL NEURAL NETWORK

Abstract

The texture of the soil has an impact on several environmental processes. Traditional soil textural assessment takes a long period, thus the quick and non-intrusive forecast of soil sand, clay, and silt seems preferable. The project's objective is to develop a hybrid Random Forest with an Artificial Neural Network (RF-ANN) model that predicts soil texture (concentrations of sand, clay, and silt) using Convolutional layers and a Random Forest. The standardized ANN model can be used to create high-resolution soil maps in parallel locations without the requirement for extra field surveys. The hybrid RF-ANN model was trained and tested using field measurements of soil texture in these land use and coverage area frames using the LUCAS Soil Texture Survey data set. Data showed that the optimization process outperformed the popular training method based on robust back-propagation. When compared to existing models, colour attributes perform better than all image-extracted kinds and have the highest influence on suggested models' performance. When compared to current models, OC, CEC, Clay, Sand, Clay, pH, and N have higher forecast accuracy. The results show that the RF-ANN model paired with linear function has been employed in the regions where the prototype is attuned if the comparative range of input parameter is similar to the section where the prototypical was regulated.

Keywords: Soil Texture, Random Forest (RF), Artificial Neural Networks (ANN), Convolutional Neural Network (CNN).

Authors

K. Anandan,
Assistant Professor
Department of Computer Applications
Nehru College of Management
Coimbatore.
ncmanandan@nehrucolleges.com

E. Chandra Blessie
Assistant Professor
Department of Computing
(Artificial Intelligence and Machine Learning)
Coimbatore Institute of Technology
Coimbatore.
blessie@cit.edu.in.

Jayamol P. James
Assistant Professor
Computer Science Department
Sahrdaya College of Advanced Studies
Kodakara , Kerala.

Copyright © 2023 Authors

I. INTRODUCTION

Soil texture is vital and exceeding variable feature of soil that affects fertility, water holding capacity, and hence production. Soil texture, which is one of the most important soil properties, affects water storage, cation exchange capacity (CEC), soil fertility, and soil ventilation. In the word "texture" indicates to the size ranges of soil particles. Skilled soil scientists can measure it in the lab or work it out in fields. When a moist soil ball (bolus) is moved between the thumb and forefinger in the field, texture is assessed as the "feel" of the soil. The quantity of sand, silt, or clay in a sample input, as well as other soil modules like organic matter or calcium carbonate, affects how moist the soil feels when massaged in the hand. Accurate and exact soil texture data are necessary for farmers who use precision-agriculture management techniques to maximize decisions regarding yield and environmental stewardship.

Soil prediction is a challenging research topic because of the various structural, physical, and chemical aspects of soil texture that make it more complex. The solid soil component, which is covered in organic matter and minerals, is reflected in the solar reflectance spectra of objects collected from minor wave solar radiation. These physicochemical parameters are thought to have the greatest influence on a soil example's reflectance when paired with sensor-measured case and reflected radiation, and are used to classify the soil prediction type. Hyperspectral sensors [2] that are sensitive to the absorption properties of organic and mineral content are used to determine it. These physicochemical factors, when paired with sensor-measured incident and reflected radiation, are assumed to have the most impact on a soil sample's reflectance and are used to identify the soil prediction type. Hyperspectral sensors [3] that are sensitive to the absorption properties of organic and mineral content are used to determine it. To use it, one must first understand functional relations among soil, other site factors, and practical field operations. Traditional methods for determining and classifying soil texture are expensive, time-consuming, tedious, and subject to human error. Recently, information on soil texture has been obtained from high-resolution satellite and aerial photos of bare soils.

In order to use remote sensing products for hydrological applications and water resource management, the scientific community has worked hard to increase the accuracy and spatiotemporal coverage of these products. Soil texture is taken into consideration while using PedoTransfer Functions (PTFs), which are built using soil components (physical, chemical, and biological) and are frequently used to assess the challenging-to-measure hydraulic soil variables. Particularly, the amount of clay in the soil has a significant impact on its water-holding capacity, which is a measure of its hydraulic qualities. It has been suggested to use a range of comparisons to evaluate soil texture modules using remotely sensed optical images. To determine soil color, these methods mainly rely on the interpretation of ground reflectance data and optical data. The use of field soil texture detection is limited across a wide area and is time- and money-consuming. High-resolution soil texture identification, which is frequently used in research, is therefore restricted to small areas. High-resolution soil texture was created using the Artificial Neural Network (ANN), Convolution Neural Network (CNN), and Random Forest (RF) models. A landscape's high-resolution soil texture, soil drainage classes, and soil organic content may all be predicted pretty well and inexpensively using ANNs and RF.

In accordance with the organizational principles of biological nerve systems, ANNs are pattern recognition tools in which the neuron or node functions as the fundamental computing element. The neuron uses two steps to process info. An internal stimulation function groups the incoming signals at first. The output of the internal activation function is taken in by a transfer function, which decides whether or not the neuron will send an output message. Neurons are linked together in layers. The strata are as follows. The distribution of soil particles—which can be categorized as clay, silt, or sand—determines the texture of the soil. Clay is defined as particles with a size less than .002 mm, silt is defined as particles with a size between .002 and .053 mm, and sand is defined as particles with a size between .053 and 2 mm. The texture of the soil affects its ability to supply water and fecundity. Soil texture and biological substance are two important factors that influence soil water retention capacity. The larger surface area of smaller particles in silt and clay allows soil to hold more water. Sand with greater particles and a minor surface area will only be able to hold a small amount of water.

For segmentation, feature extraction, and image classification to be successful and robust, data must be pre-processed. According to Boggess (1994), it is crucial to extract both latitudinal and spectral facts of information from photographs in order to utilize them. Roads and other characteristics are distinguished in Landsat imagery using ANNs. In applications like multispectral image processing, where there is no a priori knowledge of data distributions, the capabilities of ANNs for non-linear occupation estimates, data classification, non-parametric regression, and non-linear decision making are critical. Post-processing of findings from training ANNs is essential for generating reliable and consistent results when tackling challenging photo recognition and analysis problems.

In an RF classification algorithm, the generation of numerous decision trees [2] is based on (randomly) choosing only one subset amongst all offered samples in RF. Each tree chooses classes, and the one with the most votes is declared the winner. This suggests that RF is a more advanced Decision Tree Classification variation. In the context of remote sensing, the method has been categorized using multispectral data, multitemporal pictures, SAR, and optical data. Three crucial factors that affect the efficiency of the RF classifier are the number of trees (the more trees, the better the accuracy), the maximum depth of the tree (a tree cannot have a depth greater than 25), and the minimum number of samples at each node.

Deep learning is a data representation learning model with several levels of processing. Deep learning differs from traditional neural networks used in soil spectrum handling since it comprises further layers and subterranean structures. Deep neural networks can automatically find the representations required for prediction using unprocessed or raw data [3] (such as photos or spectra). At each layer, the data is changed to improve prediction by emphasizing key components of the incoming data and reducing irrelevant information. To create an RF classifier, Classification, and Regression Trees (CART) must be generated from the samples. A subset of input features, sometimes known as attributes, is chosen at random to produce a group of predictor variables. Executing a unique knowledge method that uses an attribute value test to partition the input variables into subsets allows for the construction of each tree. Every tree gives the greatest common class in each instance a unit of choice as a classification unit. The ultimate designation is decided by a popular vote of all the trees. In this work, we used the Land Use and Coverage Area Frame Survey (LUCAS)

Copyright © 2023 Authors

data set to analyze OC, CEC, CLAY, SAND, pH, and N using RF-ANN and convolutional layers to increase the accuracy of texture recognition from the soil.

II. RELATED WORKS

We look at published research in this domain that is essential to the soil texture categorization utilizing ANN, CNN, and RF [3]. Cloutis introduces geological remote sensing in 1996. When classifying soil texture, conventional methods like maximum likelihood, nearest neighbor, nearest mean, hidden Markov models, and spectral angle matching produce good results. A sample survey and extrapolation technique is used in India to create country-level soil maps at a scale of 1: 250 000. In order to categorize the soil texture classes, several sampling points were used, and a small sample set was checked in the lab for accuracy. To map soil texture, polygons [4] were employed, with each polygon indicating a mixture of two or more. There may be a great deal of ambiguity in the textural composition inside the area that a polygon labels due to intra-polygon variability. Additionally, the texture of the soil varies with depth both vertically and horizontally as well as on the surface. For any environmental simulation at certain depth ranges, soil information, such as soil texture, is required rather than pedogenetic horizons. The equal-area quadratic spline function suggested for modelling [5] the depth function of soil attribute was employed in several soil mapping studies. Several research that use regression techniques to map soil element size for various depth intervals have been reported. In this context, the current study intends to determine texture modules for six standard depth intervals utilizing regular sand, silt, and clay content of different texture classes using an equal-area quadratic spline function.

The ANN model is trained and assessed using soil texture field data in the Black Brook Watershed (BBW) in northwest New Brunswick, Canada [5]. The steps in the process of using ANNs to model soil attributes included building an ANN structure, training the ANNs, and optimizing the network. The results showed that the Levenberg-Marquardt optimization methodology worked better than the conventional preparatory method based on the strong back-propagation algorithm. For clay, there were 4.0 root mean square discrepancies between model predictions and field measurements, whereas there were 6.6 root mean square errors for sand concentrations. For clay content, the relative overall accuracy [6] (within 5% of field measurement) was 88%, while for sand content, it was 81%. On an exploratory farm in southeast New Brunswick, around 180 kilometres from the Black Brook Watershed, the trained ANN model underwent its initial calibration. The ANN model may be employed in the regions where it was calibrated or in other regions with appropriate training because the relative range of input considerations was comparable to the region where the model was calibrated.

It was chosen because ANN did not require a hypothesis and allowed for non-linear mapping [7] with little fluctuation between response and yield data (Li, 1998). Back propagation networks were trained using a method that changed the weight and bias values along a negative gradient descent. This technique was used to decrease the mean squared error (MSE) between the input and output vectors of the training set. The model's structure [8] and coefficients were altered to 'train' it using various rules and procedures. The term "epoch" refers to a single instance of prepared data with bias and weight adjustments.

The purpose of ANN optimization is to change the network structure to increase prediction accuracy [9]. It was split into two sections, and they were tasked with selecting the best input combination. One-variable, two-variable, and three-variable combination schemes should be used, with the appropriate sum of unseen layer nodes has been selected. After the number of unseen layer nodes is too little, the ANN's prediction accuracy stayed low.

Overfitting was possible when there were too many hidden layer nodes. When the hidden layer nodes were raised from 5 to 40 and the training cycles were increased from 25 to 250, the prediction performance of the ANN trained using the LM and RP approaches was evaluated. The development of ANNs [10] for soil parameters like slope, STF, SDR, and VSP utilized a mix of coarse resolution soil data (i.e., average soil drainage, sand, clay, and silt contents) and topo-hydrological data acquired from a DEM. The nodes in the output layer represented anticipated soil characteristics. For each epoch (measured values), the MSE between the system outputs (predicted values) and the battered values was determined. Training persisted even after the MSE could no longer be decreased.

For learning hyperspectral data-soil texture correlations, machine learning techniques [11] are well-suited. Machine learning techniques are divided into two categories: deep learning and shallow learning. In the past, support vector machines, random forests, and self-organizing maps all performed well with hyperspectral estimation tasks. Deep learning techniques, or network topologies with multiple hidden layers, have been the subject of recent research. Convolutional neural networks are one type of deep neural network.

III. METHODS FOR SOIL TEXTURE CLASSIFICATION

Various machine learning algorithms [12] are taught on a specific training dataset. Pre-trained networks are used to classify hyperspectral pictures via transfer learning. In transfer learning, the trained properties of a neural network are considered equivalent across different picture datasets. As a result, this strategy saves time and allows for smaller datasets to be trained. The model is made up of layers, with convolutional and assembling layers being the first. The units of the convolutional layers are organized into feature maps and connected to local patches in the feature map of the preceding layer via filter banks. The neural network is pre-trained on a certain dataset in order to enable the latter. They used the CNN for regression of the soil's clay content based on the LUCAS soil and a one-dimensional transfer learning CNN [13] as a solution. Considering that the neural network has already undergone extensive pre-training using a different dataset. Transfer learning is created specifically for 1D. Using the LUCAS soil dataset as a basis, they employ CNN for regression of the soil's clay concentration. Two-dimensional (2D) spectrograms were used to represent the raw spectrum data, and they performed better than conventional techniques like partial least squares regression (PLSR) [14]. A spectrogram is created by employing a short time quick Fourier transformation to separate a spectrum signal into overlapping components.

Modern developments to the precision farming and sensors have been resulted in gainful farming businesses. The usefulness of on-the-go soil sensors has been established in a number of researches. Advanced technologies have made it possible to obtain detailed data on the soil microclimate. Profile samples were gathered and organized as part of the Soil Resource Mapping (SRM) project (1989-1999) and other ICAR-NBSS&LUP initiatives in order to map the various soil texture classes in the research area. The morphological and

Copyright © 2023 Authors

physical properties of these soil profiles were all carefully examined, and their locations were identified through relationships between soil landforms [15]. Up to 200 cm, or until rock depth was reached, they were investigated, and every feature of their morphology and physical makeup was meticulously scrutinized. The soil summaries were divided into different soil chains based on identifying criteria (Soil Survey Staff, 2003). Samples of soil were taken from the typical master summary of the selected soil series for laboratory analysis.

The intensity change of laser beams [16] going through the soil-water mixture was collected using a microcomputer, and computerized calculations were produced using the Support Vector Regression technique. When compared to the traditional hydrometer method, the approach's success rate for texture analysis is higher. In the trials, the amplitude of light signals travelling through the soil-water mixture was determined using ten different soil samples. The results of the sand, silt, and clay estimations are compared to those of a typical hydrometer analysis. This study aims to construct a convolutional neural network classification model [18] for the soil of six distinct land cover types in Qingdao, China: orchards, forests, tea farmsteads, farmlands, bare regions, and heaths. Using the support vector machine technique, the classification results from multiple training data sets are assessed and contrasted. When the calibration set is randomly divided into portions of 1/3 and 1/4, the classification scores of the convolutional neural network likewise rise.

On 40 soil pictures, the proposed pH value prediction is tested [19]. These color models are used to calculate the pH aspect of every soil appearance. Accuracy and RMSE values are computed after applying a range of classifiers to each color space model. As a result, the main objective of the system is to predict the optimal soil pH so that the crop can be predicted using the pH data. The pH readings are computed once the soil samples have been processed. In order to determine how effectively soil attributes from two data sets and remote sensing data can predict the Mean Weight Diameter (MWD) as a measure of soil aggregate stability, the proposed research will assess machine learning algorithms (Random Forest and Multiple Linear Regression (MLR)) [20]. In terms of forecasting soil aggregate stability, the results of both models were satisfactory and comparable. As a result, the performance of the model was unaffected by the addition of remote sensing indices to soil attributes. In both representations, biological difficulty is the most crucial element in predicting soil collective stability.

The strata were collected, the percentages of clay, sand, and silt in soil samples were calculated, and 115 summaries were selected using the Latin hypercube sampling technique [21]. Researchers used ANN, Regression Trees (RT), and neuro-fuzzy (ANFIS) simulations to determine the relationship between soil records (clay, sand, and silt) and auxiliary variables. The study also revealed that spectrometric data, multi-resolution, valley-bottom flatness index, and wetness index were likely the most important auxiliary components. In general, in other arid portions of Iran, ANFIS models are recommended for numerical representing of soil texture fractions. A different method for estimating a large range of soil physical properties is reflectance spectroscopy [22]. It's a low-cost, rapid, and repeatable method of analysis. VIS-NIRS spectroscopy and the standard laboratory technique were used to assess the texture of 100 soil samples (0-20cm). The Partial Least Squares Regression (PLSR) technique was used to assess the capacity of supernatural data to forecast soil texture. The maps of projected and measured soil texture, on the other hand, showed excellent spatial

Copyright © 2023 Authors

similarity for the sand percentage, a slight difference in the variability of the clay fraction, and a smaller difference in the variability of the silt fraction.

IV. PROPOSED MODEL

1. **Hybrid RF-ANN model:** Clay, sand, and silt particles make up the soil texture. These particles play a significant role in agricultural crop production, determining yield levels. When these particles are out of equilibrium, organic matter must be supplied to the soil to artificially restore balance. By adding chemical additions, the soil's growth rate is harmed; therefore, it is preferable to use organic particles. Organic matter is also added to the soil, and making the soil more productive is not easy. Because it takes time for the organic particles to combine with the soil mixture, we must add them gradually. As a result, a prediction approach that assists may be required for crop yield and growth measurements in agriculture. Because of its nature and the hydrological process, understanding the structure of the soil is quite challenging. It is vital to predict the soil characteristics before beginning the cultivation procedure. This will aid in estimating plant growth rates and water consumption in order to balance the hydrological processes associated with soil and growth rate prediction. Once these factors have been forecasted, we will be able to analyze soil nutrients and weather conditions to determine viable crops and fertilizer requirements.

2. **RF-ANN model:** In this purpose system has been to find an efficient and reliable random forest with ANN-based soil texture classification using data from the LUCAS soil texture data set from Europe. The major goal of this study was to provide processes and methods for using RF with ANN to categorize soil texture data, as well as to compare the outcomes of RF with ANN and CNN classification. Because different parametric models give varying units of precision in changing soils and settings, applying a single parametric model frequently results in a practical connection that is more accurate in certain sections of a textural domain but less suited in others. One potential solution to this issue is the employment of the ensemble technique, which combines the best components of many parametric models to draw conclusions that are more general from fitting the data to the actual values measured. Furthermore, as demonstrated, the proposed meta-model may often address systemic problems (underestimation, overestimation, multiplicative error, etc.) caused by specific representations.

   For these reasons, the collaborative approach has been utilized to improve the generalizability and robustness of any of the constituent models. The collaborative model, which is a data-driven model based on the RF technique, uses the outputs of these parametric representations as inputs, showing that a stacking type of ensemble was utilized. In stacking, which computes the results and, for instance, replaces the average technique used in bagging, base models and meta-models are initially provided. Before deciding how to most effectively combine the output of the base models to get the desired outcome, stacking determines which base models are more trustworthy than others using the previously described meta-model. An RF approach is used to solve the learning problem, and the suggested outcomes of the base representations are actually new data for the problem.

Copyright © 2023 Authors

The RF forecast depicted in figure 1 is an average of all of these tree earnings criteria. Each regression tree in the RF approach was constructed using a bootstrap sample of the training data. The user decides how many trees to include in the RF ensemble. Each tree was trained using a sample made by randomly selecting N cases from the original dataset and replacing them with new ones, where N is the total number of variables in the original dataset. These bootstrapped prepared sets each produce a unique tree. By averaging the values predicted by each regression tree, the RF forecast is determined.

3. **LUCAS data set:** The Land Use/Land Cover Area Frame Survey (LUCAS) was expanded by the European Commission in 2009 to include topsoil quality data collection and analysis in 23 EU Member States. With all soil samples tested in a single laboratory, this topsoil survey is the first attempt in the EU to develop a consistent spatial database of soil cover based on accepted sampling and analytical procedures. About 20,000 locations from the primary LUCAS grid were selected for the collection of soil samples. To gather about 0.5 kg of topsoil, a systematic sampling approach was followed (0-20 cm). Physical and chemical evaluations of the samples were sent to a central laboratory. In total, 19,967 geo-referenced samples from 25 nations are included in the final database. The information is publicly available and can be downloaded after completing the Request Form.

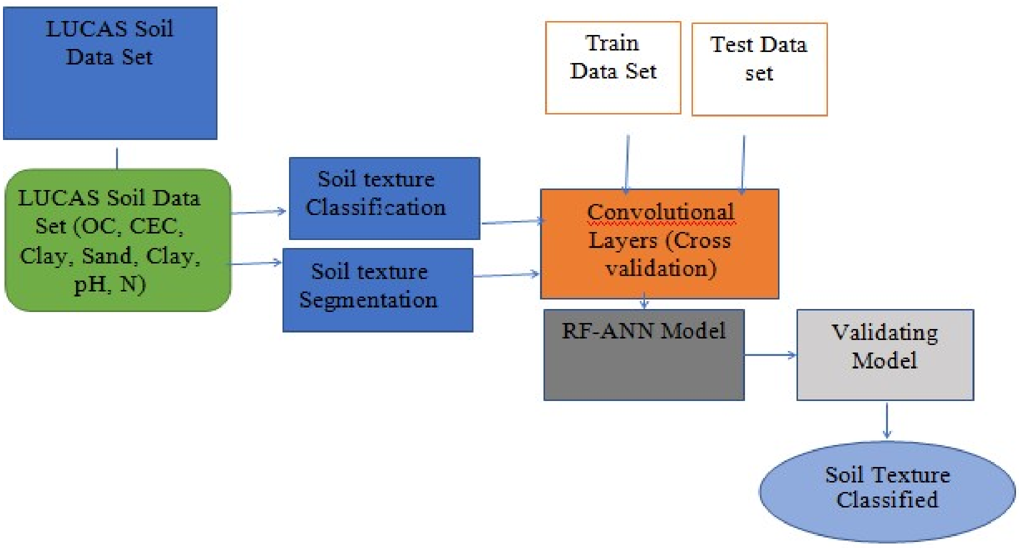

Figure 1: Design of RF-ANN model to identify soil texture

V. RESULT AND DISCUSSION

1. **Optimization results of RF-ANN Model:** Data-driven models must be optimized if they are to produce outcomes that are accurate and trustworthy. The basic goal of model optimization is to achieve the highest ideal parameters. Three important variables are included in an RF-ANN: NT (number of trees to develop), NRandam (number of variables randomly picked as candidates at each tree split), and Nsize (minimum size of

terminal nodes), which has the greatest impact on the model's final precision. In this study, an RF-ANN is tuned using grid search and repeated cross-validation.

In these network, searches is used in the optimization process to select values for both model parameter by selecting sequentially from a network of predefined values and then calculating with them. The optimum parameter combination is picked from the iteration in which the maximum level of model precision was obtained. This precision is determined by averaging the results of multiple cross-validation tests. In figure 2, we apply a technique known as repeated cross-validation. The first step in this procedure is to divide the training sample into numerous roughly equal-sized datasets called folds at random. Many folds are there to process except one used as inputs to the model during the training process, and the one-fold which is not used for authenticating data. This technique is repeated as many times as the number of folds produced by each parameter combination. The novel random dividing of the training data into folds is recurrent multiple times in repeated cross-validation. This process is repeated in order to acquire a more generalised model evaluation.

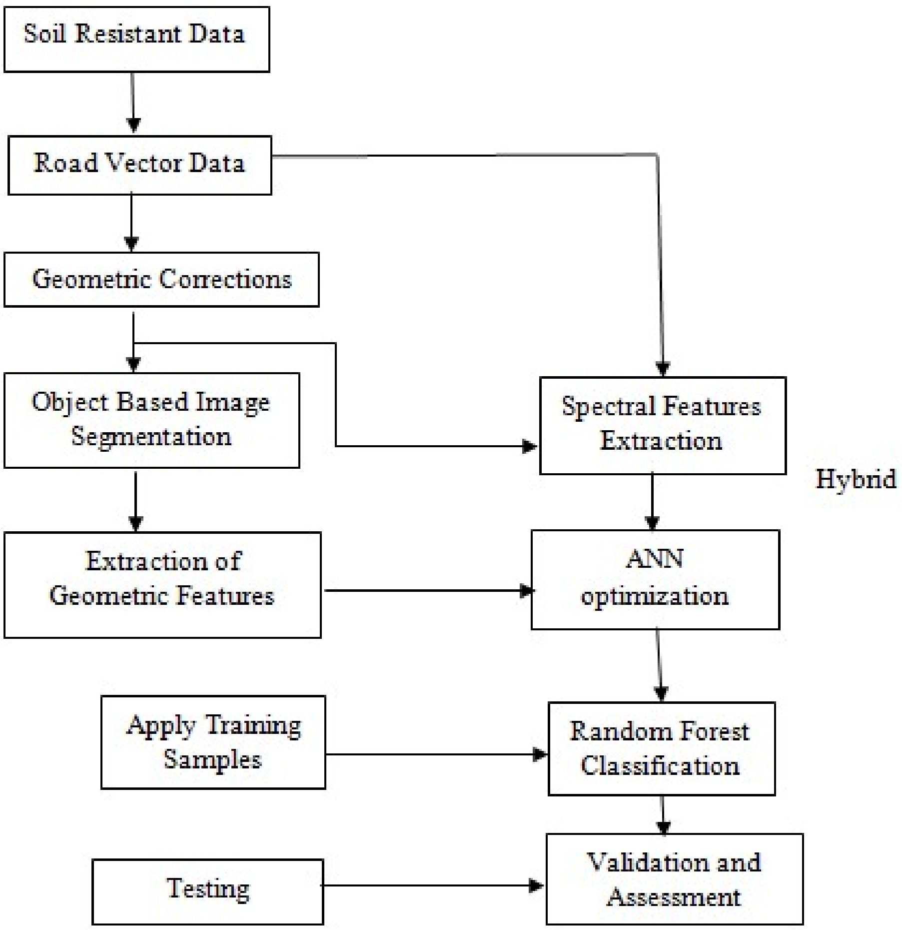

Figure 2: Architecture of Hybrid RF-ANN model

Algorithm 1: Hybrid RF-ANN Classification Algorithm for Accuracy

Input : Soil Resistant Training Group Data
Output: ANN-RF Accuracy prediction
Step 1: Taken iteration = 500
Step 2: For i = 1 to Max. No. Of Iterations
Step 3: Take soil resistant Train and load vector data.
Step 4: Perform Hybrid validation of Load group vector data.
Step 5: Create Hybrid classifier performance object, LUCAS (OC, CEC, Clay, Sand, pH, N) Vector data.
Step 6: End for

Algorithm 2: Linear Internal Function for Hybrid RF-ANN method

Step 1: Perform Hybrid RF-ANN Training to create the RF-ANN classifier using confusion matrix and Group vector data
Step 2: Now, measure the training test - Result by using group and Test vector soil data set.
Step 3: Validate the RF-ANN classifier performance object LUCAS (OC, CEC, Clay, Sand, pH, N) for test vector by considering test – result and measure the accuracy.
Step 4: Calculate the RF-ANN Accuracy of the confusion matrix in percentage.

2. **Root mean square error (RMSE):** The RMSE in tables 1 and 2 represents the sample standard deviation of the differences between anticipated and observed values. The goal is to create a single predictive power and accuracy score by combining the magnitudes of prediction mistakes. Due to its scale dependence, which is not between variables, it can only be used to evaluate the prediction errors of different models for a single variable. You can think of each RMSE as the average prediction error for a specific scale (unit).

Table 1. Training data RMSE Error

	OC_RMSE		CEC_RMSE		clay_RMSE		sand_RMSE		pH_RMSE		N_RMSE	
	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN
Mean	6.11	4.1	3.81	2.1	4.75	4.2	9.2	8.13	0.48	0.32	1.30	0.55
Min	5.39	3.42	3.47	2.35	4.23	3.15	8.69	8.12	0.67	0.59	1.22	0.42
Max	8.35	7.04	5.03	4.2	6.34	5.53	10.36	9.25	0.64	0.56	1.38	0.98

Table 2. Test data RMSE

	OC_RMSE		CEC_RMSE		clay_RMSE		sand_RMSE		pH_RMSE		N_RMSE	
	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN
Mean	8.83	7.68	8.68	4.49	7.47	7.62	12.03	11.11	0.50	0.81	1.52	0.7
Min	6.93	5.16	7.97	2.81	7.06	6.37	8.21	9.09	0.62	0.57	1.47	0.51
Max	10.86	8.73	9.65	5.62	7.78	8.8	15.01	14.97	0.52	0.45	1.59	0.84

Copyright © 2023 Authors

Table 3. Training data Accuracy

	OC_Acc		CEC_Acc		clay_Acc		sand_Acc		pH_Acc		N_Acc	
	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN
Mean	0.85	0.89	0.84	0.86	0.82	0.89	0.86	0.92	0.9	0.95	0.66	0.75
Min	0.93	0.90	0.69	0.63	0.7	0.82	0.8	0.86	0.8	0.85	0.06	-0.09
Max	0.98	0.97	0.88	0.90	0.85	0.89	0.89	0.94	0.94	0.96	0.87	0.92

Table 4. Test data Accuracy

	OC_Acc		CEC_Acc		clay_Acc		sand_Acc		pH_Acc		N_Acc	
	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN	CNN	RF-ANN
mean	0.89	0.91	0.71	0.79	0.72	0.78	0.62	0.75	0.86	0.89	0.85	0.90
Min	0.84	0.88	0.65	0.72	0.67	0.72	0.6	0.74	0.87	0.87	0.82	0.85
Max	0.86	0.82	0.83	0.89	0.77	0.85	0.85	0.89	0.88	0.90	0.87	0.92

The data is presented as conditions, each cell containing a count. The counts of the anticipated class values are displayed vertically, while the sums of the actual class values are shown in parallel. A perfect set of predictions can be represented by a diagonal line running from the top left to the bottom right of the matrix. For classification issues, a confusion matrix is helpful since it identifies which predictions were wrong and what kind of error was made. Writing a purpose to evaluate the confusion matrix based on a list of actual class values and forecasts can be the first step. Each class value in the confusion matrix has been given a distinct number or index, and a list of all the unique class values has been created. The confusion matrix, which is always square, must have the specified number of rows and columns based on the number of class values.

Equation 1 evaluates a collection of predictions on a classification job using accuracy. The classification accuracy in tables 3 and 4 is calculated as the number of correct answers divided by the total number of predictions. Usually, it is stated as a percentage between 0% and 100%, with 0% denoting the best possible accuracy and 100% denoting the worst possible accuracy.

Accuracy = correct predictions / total predictions * 100 ---------------- (1)

This can be accomplished by using a function that takes forecasts and expected results as parameters. By averaging the evaluated statistic (such as root mean square error, or RMSE) from all model runs with a specific set of parameters, this method calculates the model's precision. The statistic that results, for instance, from two repetitions and five folds is the mean of the 10 numbers. The tuning notion in this work serves two purposes: determining the appropriate RF-ANN algorithm parameters and estimating the suggested model's precision for upcoming data.

Copyright © 2023 Authors

VI. CONCLUSION

The proposed deep learning-based technique for predicting soil texture for better understanding of yield prediction, using OC, CEC, Clay, Sand, Clay, pH, and N. To develop hybrid RF-ANN, the proposed model is a hybrid that combines convolutional layers with multiple filters utilizing Random Forest and ANN. The model's ANN component is designed to capture the intrinsic temporal and spatial correlations of meteorological and soil data collected at various depths under the surface. As a result of continual developments in plant breeding and management practices, the CNN layer of the model was constructed to reflect the criteria to determine the optimal soil for developing trends in crop output over time. The neural network model is made up of convolutional layers and hybrid RF and ANN layers, with the network performing preprocessing and regression prediction of six soil textures. Raising agricultural productivity is a critical step toward eliminating poverty in the future, but farmers face considerable challenges in increasing output. Poor soil, inefficient water use, and a shortage of plant breeding resources, nutritious animal feed, and high-quality seed are to solve these issues.

REFERENCE

[1] John K, Abraham Isong, Michael Kebonye N, kon Ayito E, Chapman Agyeman P, Marcus Afu S, Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil. *Land* 2020, *9*, 487. https://doi.org/10.3390/land9120487

[2] FF. M. Riese and S. Keller, Soil Texture Classification With 1d Convolutional Neural Networks Based On Hyperspectral Data https://doi.org/10.5194/isprs-annals-IV-2-W5-615-2019.

[3] Omosalewa Odebiri, John Odindi Onisimo, Mutanga, Basic and deep learning models in remote sensing of soil organic carbon estimation. https://doi.org/10.1016/j.jag.2021.102389

[4] Antonios Morellos, Xanthoula, Eirini Pantazi, Dimitrios Moshous, Thomas Alexandridis, Rebecca Whetton, Georgios Tziotzioz, Jens Wieben sohn, Ral Bill, Abdul M.Mouazen, Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using vis-nir spectroscopy. https://doi.org/10.1016/j.biosystemseng.2016.04.018

[5] R.K.Swetha, Prawal Bende, Kabeer Singh, Srikanth Gorthi, Asim Biswas, BinLi, David C.Weindorf, Somsubhra Chakraborty, Predicting soil testure from smartphone captured digital images and an application. https://doi.org/10.1016/j.geoderma.2020.114562

[6] K Anandan, R Shankar, S Duraisamy, Convolutional Neural Network approach for the prediction of soil texture properties. DOI:10.17485/IJST/v14i3.2047

[7] Abraham, S.; Huynh, C.; Vu, H. Classification of Soils into Hydrologic Groups Using Machine Learning. *Data* **2020**, *5*, 2. https://doi.org/10.3390/data5010002

[8] Freddy Bangelesa, Elhadi Adam , asper Knight, Inos Dhau, Marubini Ramudzuli and Thabiso M. Mokotjomela, Predicting Soil Organic Carbon Content Using Hyperspectral Remote Sensing in a Degraded Mountain Landscape in Lesotho, https://doi.org/10.1155/2020/2158573

[9] Wartining, Budiman Miasny, Maryam Montazerolghaem, Jose Padarian, Richard Ferguson, S.carlett Bailey, Alex B.McMratney, Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. https://doi.org/10.1016/j.geoderma.2019.06.016

[10] B. Zhao, B. Huang and Y. Zhong, "Transfer Learning with Fully Pretrained Deep Convolution Networks for Land-Use Classification," in *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1436-1440, Sept. 2017, doi: 10.1109/LGRS.2017.2691013.

[11] Maher Ibrahim Sameena, Biswajeet Pradhana,b, Saro Leec.D, Application of convolutional neural networks featuring Bayesian optimization for landslide susceptibility assessment. DOI:10.22055/AAG.2019.29642.1990

[12] Petersson, H., Gustafsson, D., Bergstr, D., 2016. Hyperspectral image analysis using deep learning-A review. In Proceedings of the 2016 6th International Conference on Image Processing Theory Tools and Applications (IPTA), Oulu, Finland. pp. 1-6. doi:10.1109/IPTA.2016.7820963

Copyright © 2023 Authors

[13] Lucas Benedeta, Wilson Missina Fariaa, Sérgio Henrique Godinho Silvaa, Marcelo Mancinia, José Alexandre Melo Demattê, Luiz Roberto Guimarães Guilhermea, Nilton Curi, Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy, https://doi.org/10.1016/j.geoderma.2020.114553

[14] O'Rourke, S.M., Minasny, B., Holden, N.M., McBratney, A.B., 2016b. Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. Soil Sci. Soc. Am. J. 80 (4), 888–899. https://doi.org/10.2136/sssaj2015.10.0361.

[15] Cisty, Milan, Celar, Lubomir, and Minaric, Peter, Conversion Between Soil Texture Classification Systems Using the Random Forest Algorithm. https://doi.org/10.1177/ASWR.S31924

[16] Yassine Bouslihim, Aicha Rochdi and Namira El Amrani Paaza, "Machine learning approaches for the prediction of soil aggregate stability", Volume 7, No. 3, March 01, 2021.

[17] Khanbabakhani Lnaz, Mohammadi Torkashvand Ali, Mahmoodi Ali Mohammad, "Prediction of Soil Texture Using Artificial Neural Networks", Journal Of Soil And Water Resources Conservation, 2018, Volume8,Number1,pp.1-9https://www.sid.ir/en/Journal/ViewPaper.aspx?ID=692940

[18] .Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. and Fernández-Ugalde, O. (2018), LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur J Soil Sci, 69: 140-153. https://doi.org/10.1111/ejss.12499

[19] E. Kılınç, F. Albayrak and U. Orhan, "Soil Texture Prediction With Laser Guided Bouyoucos and Support Vector Regression," 2020 Innovations in Intelligent Systems and Applications Conference (ASYU), 2020, pp. 1-4, doi: 10.1109/ASYU50717.2020.9259844.

[20] Xueying Li, Pingping Fan, Zongmin Li, Guangyuan Chen, Huimin Qiu, Guangli Hou, "Soil Classification Based on Deep Learning Algorithm and Visible Near-Infrared Spectroscopy", Journal of Spectroscopy, vol. 2021, Article ID 1508267, 11 pages, 2021. https://doi.org/10.1155/2021/1508267

[21] Wani T., Dhas N., Sasane S., Nikam K., Abin D. (2021) Soil pH Prediction Using Machine Learning Classifiers and Color Spaces. In: Joshi A., Khosravy M., Gupta N. (eds) Machine Learning for Predictive Analysis. Lecture Notes in Networks and Systems, vol 141, 2020, Springer, Singapore. https://doi.org/10.1007/978-981-15-7106-0_10

[22] Mehrabi-Gohari, E.; Matinfar, H.R.; Jafari, A.; Taghizadeh-Mehrjardi, R.; Triantafilis, J. The Spatial Prediction of Soil Texture Fractions in Arid Regions of Iran. Soil Syst. 2019, 3, 65. https://doi.org/10.3390/soilsystems3040065.

Copyright © 2023 Authors