Soil Organic Carbon Stock Predictability based on Multispectral Remote Sensing Data - A Review

Susanta Kundu[1], Vinod Kumar Sharma[2]

[1,2]Department of Computer Science and Engineering, Shree Guru Gobind Singh Tricentenary University, Gurugram

Abstract

Soil Organic Carbon (SOC), usually located within a meter of the topsoil layer, is the largest terrestrial carbon pool. Rapid urbanization and industrial growth have drawn attention to SOC Stock (SOCS) in soil, equating its correlation with carbon emissions resulting in global warming in terrestrial environments. Such transformed global warming correlates with initial soil organic carbon stock size. Estimating soil organic carbon stock may help to limit the earth system model projections in global warming [1].

This paper reviews commonly used predictive statistical models to forecast SOC estimation using popular algorithms in machine learning and deep learning. Input data used to 'train' by different algorithms can predict SOCS with variances in their performance. This study covers relative comparisons of Artificial Neural networks (ANN), supporting vector machines (SVM), and other Deep Learning models used in computing SOCS with multispectral remote sensing data. In general, comparing and choosing the optimal technique is an exciting task.

Keywords: Soil Organic Matter, Soil Organic Carbon, SVM, ANN, Machine Learning, Deep Learning

1. Introduction

Carbon sequestration is the potential in mitigating climate [1] and was highlighted at COP21 – the 21st conference in the United Nations (UN) Convention on Climate Change. Soil organic matter comprises microbial biomass dissolved and humas from organic substances. These substances dissociate into nutrients useful for vegetation and plant growth, adding essential ingredients to soil and its structural health [2]. The extent of soil organic carbon stored (SOCS) in soil may vary on the adjoining landscape, local climate (temperature, rainfall, and precipitation), land inclination /slope, farming methodologies adopted for soil nutrition, etc. [3].

SOCS measures soil fertility by relating its richness to growing vegetation and agricultural products, including crops [4]. Soil has the most extensive stock of organic carbon on the earth, covering nearly ¾ of the total carbon reserve of the earthly ecosystem, interacting closely with the climatic, atmospheric, and land cover changes [5]. SOCS estimation is important at a local, regional and global level due to its influence on agriculture production, climate control, and global warming mitigation [6].

SOC, usually stored in the soil at a 20 – 30 cm depth, is key as a terrestrial carbon pool deposit [7]. Over time, SOCS depletion with land-use changes is a worldwide environmental issue [8].

Urbanization and extensive human usage of sub-soil matters release stored carbon in the soil, releasing carbon dioxide into the atmosphere [9]. Soil samples from 0–20 cm soil depths are more accessible to collect as soil samples for testing in the laboratory to measure SOC and derived SOCS. Soil samples are tested using 400–700 nm wavelength spectroscopic methods at the designated study locations. The soil reflectance spectra of these study areas, as measured by appropriate satellite bands, are helpful to evaluate the performance precision of related trained models [10].

An estimated SOCS increase of 0.4% per year is considered adequate to reduce Green House Gas emissions [11]. SOCS affects the soil characterization through its physical, chemical, and biological properties, thereby enhancing water and nutrient retention in soil [12]. Thus, consistent and reliable data on SOCS is essential for the systematic monitoring of SOC and spatial mapping capacity [2].

2. Related works in Soil Organic Carbon Stock (SOCS)

Importance of SOC: (Climate – Greenhouse warming – Soil Health – Agricultural benefit)

SOC content in soil is dynamic in space and time. Such dynamism is related to carbon sequestration, land usage, degradation due to human interventions, and climatic conditions [6]. An improved understanding of SOCS conditions spatially and time series is indispensable for climate change mitigation options. Managing and parameterizing SOCS is essential for an evolved climate control policy globally impacting the carbon cycle. Urbanization, changes in Land Use (LU), harmful agricultural practices, etc., play crucial factors in SOC changes [13].

Regular SoC estimation and monitoring are necessary over spatial and time domains in a specific landscape scale to understand the overall dynamics of SOC changes [7]. Regular and updated information on land use and SOCS in soils, monitoring, and mapping of SOCS changes in regional areas are essential to improve decisions on land degradation management and policy developments [8].

A data-centric statistical model to map SOCS in space and time is needed for an annual SOCS prediction considering the dynamism in SOC content. The process requires a SOCS capacity map from a reference year as a baseline [14]. Many countries across the globe have commonly adopted this methodology to report carbon trends to the UNCCD (IPCC, 2006; Mattina et al., 2018).

SOCS monitoring needs to maintain SOC data for an appropriate target to mitigate climate changes. Existing global SOC maps have limited information and coverage at a regional scale to make decisions to track carbon sequestration on a time scale. The prediction map as an estimation will need a regular or timely update of SOC measurement to monitor SOC changes and how it is supplementing to ease climatic issues [15].

The land-use changes show that converting forestland and wetland to cropland caused SOCS loss from the soil. Such changes challenge our ecosystem, adversely impacting SOCS [16].
The predictability of SOCS changes requires technological advancements in collecting and processing satellite-based remote sensing data [17]. It also involves soil data

from different regional and global databases in published documents or databases to supplement SOCS predictability analysis [5].

Soil samples collected from study plot areas are primary field data elements tagged with positional information (latitude, longitude) for georeferencing. Under standard laboratory procedures, the soil samples were examined for SOC-related tests using total oxidized carbon and wet oxidation measurement as per standard laboratory procedures [18].

3. SoC Estimation using Multispectral Remote Sensing Data

The traditional field SOC sample collections followed by laboratory testing are relatively expensive and time-consuming [21]. Such limitations encouraged researchers to explore using geographical information systems (GIS) and remote sensing (RS) techniques to map and monitor land uses [22].

Soil reflectance spectroscopy using sensors mounted on satellites has gained momentum for soil organic carbon estimation in various research areas, with the methodologies and reported results published in each study [20]. The current trend to source RS data to evaluate SOC and SOCS is a cost-effective, readily available, and non-destructive approach for SOCS estimation [19]. It, however, still uses the field data as reference data to calibrate and benchmark the SoC estimation model [23].

Widely available spatial and time domain datasets and advanced computer vision through deep learning and other ML techniques eased the satellite image translation and related applications [24]. Advancements in deep learning (DL) algorithms have rapidly increased the popularity of RS image analysis in the last few years. Major DL development related to the remote-sensing data study targets address the spatial image resolution, study area type, and classification accuracy achieved. DL applied for remote sensing image analysis of SOC includes segmentation, classification, and image analysis of land use and land cover (LULC) [25].

The advent of multispectral satellite data availability as a source of spectral data and technically advanced data processing in deep learning algorithms made it a cost-effective, technically viable process to estimate soil organic carbon (SOC) in spatial coverage of SOC maps [26].

Satellite images from a study area can facilitate connecting SOC variation with time [8]. The acquisition of satellite data at regular intervals opened up the option of analyzing SOC contents in time series [18]. ML algorithms like Random Forest, Support Vector Machine (SVM), and Boosted Regression Test (BRT) modeled the spatial variability of SOCS with maximum likelihood and supervised classifications [8].

For remote sensing data, satellite images from the multispectral spectral bands of similar wavelengths are used for random forest regression in the SOC estimation and SOCS prediction. [10]. Remote sensing (RS) data is available from different satellites for SOC mapping in spatial resolution and time domain with varied spectral resolution. Inaccurate SOC estimations may occur in the case of the lower spectral resolution of satellite sensors bands in the visible and SWIR spectrum relevant to organic matter absorption [27]. Spectral data from these bands are potential indicators in the spatial mapping of SOC. Unlike a spectral resolution, a spatial resolution has less impact on predictions with dynamically changing soil characteristics [18].

Digital Soil Mapping (DSM) methods are part of SOC predictive models that use RS data of easy data availability and extensive spatial and time-series distribution. The RS variables widely used are surface reflectance, vegetation index, and band ratios relating to vegetation and soil [28].

In SOC mapping, band ratios or surface reflectance are effective parameters in bare soils. Connections between vegetation and SOC indicate RS data produces enhanced vegetation growth status over a longer time duration [29]. In vegetative areas, Normalized Differences Vegetation Index (NDVI) is helpful [30], though it may not enhance mapping accuracy significantly [31]. The other Multispectral RS data relevant to SOC mapping available for spectral and spatial resolution mapping and band combinations are Brightness Index (BI) and Greenness Index (GI), as covariates in SOC estimation [32]. The surface reflectance spectral data from satellite bands are considered independent variables in models evaluating the influence of spectral data to predict SOCS [29]. A list of other helpful predictor variables includes wetness index (WI), vegetation temperature condition index (VTCI), and compound topographic index (CTI) [33].

4. Machine Learning Applications for Predicting SoCS

Remote Sensing data from satellite images supplements estimating SoC Stock (SOCS) trained models. The performance accuracy of predictive statistical models, such as random forest, partial least square regression, etc., can be likened to assessing the SOC content. Predicting soil organic carbon in spatial distribution for land management relates to carbon emission and soil health [34]. Performance comparison of these models suggests that the predictability is not entirely dependent on the spectral behavior [35].

Different algorithms, extrapolation techniques such as Artificial Neural Networks (ANN), decision trees, Support Vector Machines, and other multivariate regression techniques are applied for statistical predictions. The performance comparisons are set by parameters like the Coefficient of Determination (R2), Root Mean Square Error of Cross-Validation (RMSECV) tests, and Ratio of Performance to Deviation (RPD) [36].

ML involves a process to learn from data consisting of a set of instances described by a group of features or variables. Upon completing the learning process, the training model is applied to a collection of new testing data to classify, predict, or cluster [24].

SOC prediction using a Convolutional Neural Network (CNN) is encouraging compared with Random Forest (RF) with multiple environmental factors. The results show improved accuracy of CNN with added land surface variates and other environment variables. CNN had higher prediction performance than RF regardless of the variables added, indicating CNN is a good soil mapping approach in a regional setup [29].

ML tasks are organized into supervised or unsupervised models, depending on the learning type or the learning models employed to implement classification, regression, clustering, and dimensionality reduction [37]. In case inputs are partially available and some outputs are partly missing in a dynamic environment, it constitutes reinforcement learning. Training the models, testing, and validating final predictions results for all input spectral data remain the same. The trained, supervised model can predict the outcomes (labels) of a few missing test data. For unsupervised learning, no separate test or training data is labeled. The input data is processed to identify a pattern [37]. The training data is presented as input in supervised learning while the corresponding output constructs a rule mapping input to output [38].

5. Prediction Models

Regression constitutes a supervised learning model to provide output prediction that varies according to the input variables. Standard predictive models are linear regression, multiple linear regression, and least squares regression.

Linear regression models correlate the distribution of dependent variables using a linear combination of independent variables (predictors). SOCS is one primary dependent variable for SOC estimation, with local climate, terrain conditions, and spectral parameters independent variables, depending on the study scenario. [39]

Multiple Linear Regression (MLR) is a statistical technique using multiple variables to predict a responsive outcome (predict) of a dependent variable based on two or more independent variables. It is an extension of Linear Regression of one explanatory variable. [34]

Partial Least Squares Regression (PLSR) defines linear relations among the variables, although the connections are not always linear. It helps to correlate soil organic carbon stocks and spectral information with statistically derived spectral characteristics. The mapping of PLSR models built on a local field study to a regional level is often limited due to wider spatial variation in SOC contents. [40]

Support Vector Machine (SVM) is a supervised ML algorithm used in SOCS estimation for related classification and regression. SVM tries to find a hyperplane to classify the data points distinctly in n-dimensions [25]. The kernel scale and corresponding penalty are adjusted in SVM modeling to convert training data in a non-linear decision surface to a linear equation of higher orders. SVM model performance is good in soil mapping [39].

Random Forest (RF) is a Supervised Machine Learning model used in regression and classification problems. In RF, the decision trees relate probable predictors with the target variables. Decision trees are built into RF to ensure a stable model where every tree is trained individually with distinct training data [41]. In RF prediction, it averages the results from all trees individually trained. RF trains faster than decision trees, working only on a subset of features [42]. Clustering is an unsupervised learning model for natural data grouping – for example, the k-means technique [43].

Prediction accuracy depends on various factors, like soil condition, variability of input characteristics, surface area with varying moisture, roughness due to vegetation and crop, and atmospheric conditions during image acquisition. [44]

The precision of prediction models based on data from multispectral bands of different satellites was recorded as RPD between 1.4

to 3.1 and RPIQ between 1.8 to 2.1 [18]. Prediction accuracy is obtained by ML/DL model or techniques or approaches using the coefficient of determination (R2), Ratio of Performance to Deviation (RPD), and Root Mean Square Error (RMSE) [45].

RMSEP (Root Mean Square Error of Prediction) measures actual instances linked to reference values, indicating an estimate of performance deterioration.

Table 1: Machine Learning Models For SoC Forecasting as published with scope

Training Model	Test RMSEP (%)	Test R^2p	RPD	Reference
PLS	0.27	0.66	1.45	[18]
CB	0.26	0.68	1.52	[18]
CB	0.28	0.56	1.4	[18]
SVM	0.28	0.65	1.41	[18]
CB	0.26	0.66	1.52	[18]
RF	0.24	0.74	1.65	[18]
RF	0.27	0.59	1.46	[18]
BPN	0.53	0.86	2.84	[45]
BPN	0.61	0.83	2.46	[45]
MLP	0.55	0.86	2.70	[45]
MLP	0.50	0.88	2.98	[45]
LeNet5	0.53	0.87	2.90	[45]
LeNet5	0.54	0.86	2.79	[45]
DenseNet10	0.49	0.88	3.13	[45]
DenseNet10	0.49	0.89	3.053	[45]
ML	0.546	0.348		[39]
RF	0.528	0.39		[39]
SVM	0.551	0.341		[39]
SGB	0.539	0.367		[39]
RF	0.768	0.83		[46]
SGB	0.782	0.8		[46]
SVM	0.792	0.79		[46]
ANN	0.918	0.76		[46]
MLR	0.972	0.75		[46]
PLSR	0.995	0.75		[46]
BRT*	0.497	0.266		[47]
RF	0.502	0.252		[47]
SVM	0.488	0.285		[47]
BPNN	0.56			[48]
FPNN	0.68			[48]
MLPNN	0.72			[48]
GRNN	0.51			[48]
PLS		0.792		[49]
RF		0.808		[49]
CNN		0.878		[49]
PLSR	0.73	0.51	1.43	[50]
PLSR	1.23	0.56	1.51	[36]

6. Challenges

RMSE (Root Mean Square Error) measures test/train data error with cross-validation using split data [44].

Prediction accuracy is obtained by ML/DL model or techniques or approaches using the coefficient of determination (R2), Ratio of Performance to Deviation (RPD), and Root Mean Square Error (RMSE) [45].

The precision of prediction models based on data from multispectral bands of different satellites was recorded as RPD between 1.16 to 1.65 [18].

A well-split data portion results in a reasonable assessment of the model using a trained dataset executed for unknown instances. Prediction accuracy depends on various factors, like soil condition, variability of input characteristics, surface area with varying moisture, roughness due to vegetation and crop, and atmospheric conditions during image acquisition [33].

7. Discussion

Machine learning models like SVM, RF, and SGB work better (more accurately) than MLR when compared with RSME-CV (cross-validation) of each model for model performance assessment [51].

The regression models in ML, like Support Vector Machine (SVM), Multiple Linear Regression (MLR), Random Forest Regression (RFR), and Stochastic Gradient Boosting (SGB), are tested to set the training model [41]. It is then compared with independent soil samples to validate the model in an extrapolated area. The performance comparison of these models indicates ML techniques have improved results compared to MLR and RFR, thereby providing better precision in most instances [52].

MLR performs low to correlate non-linear relations between dependent and independent variables while predicting soil parameters on unsampled positions [39]. The predictive value performance comparisons are set by parameters like the Coefficient of Determination (R2), Root Mean Square Error of Cross-Validation (RMSECV) tests, and Ratio of Performance to Deviation (RPD) [36]. The prediction performances vary among ML algorithms and models, with few performing reliably and maintaining enhanced accuracy in SOC prediction (R2 = 0.470, RMSE = 0.437 [47].

Deep Learning models produce good results in RS-based SOC mapping, despite a few limitations like acquired soil samples, laboratory tests through laborious, expensive computational time, data acquisition with cloud coverage, and inclination for over-fitting models [53]. The observations on SOCS predictions resulting in inaccuracy or inconsistency can be due to the dynamism involved in SOC values on varying local factors [54].

8. Conclusion

SOC mapping and prediction using MS RS data work through a machine learning framework. Researchers explore improved models like deep learning and RF [55]. Regression models and machine learning improve accuracy [56] with advanced theory and well-developed software applications. A combination of RS with field data through laboratory tests provides a practical approach [14].

RF model emerged as the promising variable identifiers to predict SOCS and build a model for extrapolation and generating the digital SOC map [57]. Its accuracy is assessed by the corresponding RMSE and coefficient of determination (R2) for comparison with peer algorithms. However RF model requires special attention to its regression tree development [58].

Irrespective of the advantages of RF models in SOCS prediction, there are few constrictions in respective regression tree development. RF models that underestimate the SOC high values in a training dataset can be fixed by exploiting supplementary datasets [46].

9. References:

[1] T. W. Crowther *et al.*, “Quantifying global soil carbon losses in response to warming,” *Nature*, vol. 540, no. 7631, pp. 104–108, Dec. 2016, doi: 10.1038/nature20150.

[2] T. Angelopoulou, N. Tziolas, A. Balafoutis, G. Zalidis, and D. Bochtis, “Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review,” *Remote Sensing*, vol. 11, no. 6, p. 676, Mar. 2019, doi: 10.3390/rs11060676.

[3] P. Kumar, H. Sajjad, B. R. Tripathy, R. Ahmed, and V. P. Mandal, “Prediction of spatial soil organic carbon distribution using Sentinel-2A and field inventory data in Sariska Tiger Reserve,” *Nat Hazards*, vol. 90, no. 2, pp. 693–704, Jan. 2018, doi: 10.1007/s11069-017-3062-5.
[4] E. Tola *et al.*, “Determining soil organic carbon concentration in agricultural fields using a handheld spectroradiometer: Implication for soil fertility measurement,” *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 6, pp. 13–19, 2018, doi: 10.25165/j.ijabe.20181106.4061.
[5] E. G. Jobbágy and R. B. Jackson, “THE VERTICAL DISTRIBUTION OF SOIL ORGANIC CARBON AND ITS RELATION TO CLIMATE AND VEGETATION,” *Ecological Applications*, vol. 10, no. 2, pp. 423–436, Apr. 2000, doi: 10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.
[6] J. Navarro-Pedreño, M. B. Almendro-Candel, and A. A. Zorpas, “The Increase of Soil Organic Matter Reduces Global Warming, Myth or Reality?,” *Sci*, vol. 3, no. 1, p. 18, Mar. 2021, doi: 10.3390/sci3010018.
[7] J. P. Scharlemann, E. V. Tanner, R. Hiederer, and V. Kapos, “Global soil carbon: understanding and managing the largest terrestrial carbon pool,” *Carbon Management*, vol. 5, no. 1, pp. 81–91, Feb. 2014, doi: 10.4155/cmt.13.77.
[8] K. Nabiollahi, S. Shahlaee, S. Zahedi, R. Taghizadeh-Mehrjardi, R. Kerry, and T. Scholten, “Land Use and Soil Organic Carbon Stocks—Change Detection over Time Using Digital Soil Assessment: A Case Study from Kamyaran Region, Iran (1988–2018),” *Agronomy*, vol. 11, no. 3, p. 597, Mar. 2021, doi: 10.3390/agronomy11030597.
[9] S. Kumar, Shwetank, and K. Jain, “A Multi-Temporal Landsat Data Analysis for Land-use/Land-cover Change in Haridwar Region using Remote Sensing Techniques,” *Procedia Computer Science*, vol. 171, pp. 1184–1193, 2020, doi: 10.1016/j.procs.2020.04.127.
[10] F. Bangelesa, E. Adam, J. Knight, I. Dhau, M. Ramudzuli, and T. M. Mokotjomela, “Predicting Soil Organic Carbon Content Using Hyperspectral Remote Sensing in a Degraded Mountain Landscape in Lesotho,” *Applied and Environmental Soil Science*, vol. 2020, pp. 1–11, Apr. 2020, doi: 10.1155/2020/2158573.
[11] K. Were, D. T. Bui, Ø. B. Dick, and B. R. Singh, “A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape,” *Ecological Indicators*, vol. 52, pp. 394–403, May 2015, doi: 10.1016/j.ecolind.2014.12.028.
[12] R. J. Zomer, D. A. Bossio, R. Sommer, and L. V. Verchot, “Global Sequestration Potential of Increased Organic Carbon in Cropland Soils,” *Sci Rep*, vol. 7, no. 1, p. 15554, Dec. 2017, doi: 10.1038/s41598-017-15794-8.
[13] F. Magdoff and H. Van Es, *Building soils for better crops: ecological management for healthy soils*, Fourth edition. College Park: Sustainable Agriculture Research & Education, 2021.
[14] G. B. M. Heuvelink *et al.*, “Machine learning in space and time for modelling soil organic carbon change,” *Eur J Soil Sci*, vol. 72, no. 4, pp. 1607–1623, Jul. 2021, doi: 10.1111/ejss.12998.
[15] Z. S. Venter, H.-J. Hawkins, M. D. Cramer, and A. J. Mills, “Mapping soil

organic carbon stocks and trends with satellite-driven high resolution maps over South Africa," *Science of The Total Environment*, vol. 771, p. 145384, Jun. 2021, doi: 10.1016/j.scitotenv.2021.145384.

[16] A. Suleymanov, I. Gabbasova, R. Suleymanov, E. Abakumov, V. Polyakov, and P. Liebelt, "Mapping soil organic carbon under erosion processes using remote sensing," *HunGeoBull*, vol. 70, no. 1, pp. 49–64, Apr. 2021, doi: 10.15201/hungeobull.70.1.4.

[17] L. Poggio *et al.*, "SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty," *SOIL*, vol. 7, no. 1, pp. 217–240, Jun. 2021, doi: 10.5194/soil-7-217-2021.

[18] D. Žížala, R. Minařík, and T. Zádorová, "Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions," *Remote Sensing*, vol. 11, no. 24, p. 2947, Dec. 2019, doi: 10.3390/rs11242947.

[19] Y. Xu, S. E. Smith, S. Grunwald, A. Abd-Elrahman, and S. P. Wani, "Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 123, pp. 1–19, Jan. 2017, doi: 10.1016/j.isprsjprs.2016.11.001.

[20] P. T. Sorenson, S. J. Shirtliffe, and A. K. Bedard-Haughn, "Predictive soil mapping using historic bare soil composite imagery and legacy soil survey data," *Geoderma*, vol. 401, p. 115316, Nov. 2021, doi: 10.1016/j.geoderma.2021.115316.

[21] S. Pal and Sk. Ziaul, "Detection of land use and land cover change and land surface temperature in English Bazar urban centre," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 20, no. 1, pp. 125–145, Jun. 2017, doi: 10.1016/j.ejrs.2016.11.003.

[22] A. Gholizadeh *et al.*, "Soil organic carbon estimation using VNIR–SWIR spectroscopy: The effect of multiple sensors and scanning conditions," *Soil and Tillage Research*, vol. 211, p. 105017, Jul. 2021, doi: 10.1016/j.still.2021.105017.

[23] A. Bhar, B. Feddersen, R. Malone, and R. Kumar, "Agriculture Model Comparison Framework and MyGeoHub Hosting: Case of Soil Nitrogen," *Inventions*, vol. 6, no. 2, p. 25, Mar. 2021, doi: 10.3390/inventions6020025.

[24] S. P. Mohanty *et al.*, "Deep Learning for Understanding Satellite Imagery: An Experimental Survey," *Front. Artif. Intell.*, vol. 3, p. 534696, Nov. 2020, doi: 10.3389/frai.2020.534696.

[25] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 152, pp. 166–177, Jun. 2019, doi: 10.1016/j.isprsjprs.2019.04.015.

[26] A. Gholizadeh, L. Borůvka, M. Saberioon, and R. Vašát, "Visible, Near-Infrared, and Mid-Infrared Spectroscopy Applications for Soil Assessment with Emphasis on Soil Organic Matter Content and Quality: State-of-the-Art and Key Issues," *Appl Spectrosc*, vol. 67, no. 12, pp. 1349–1362, Dec. 2013, doi: 10.1366/13-07288.

[27] E. A. Thaler, I. J. Larsen, and Q. Yu, "A New Index for Remote Sensing of Soil Organic Carbon Based Solely on Visible Wavelengths," *Soil Sci. Soc. Am. J.*, vol. 83, no. 5, pp. 1443–1450, Sep. 2019, doi: 10.2136/sssaj2018.09.0318.

[28] A. B. McBratney, M. L. Mendonça Santos, and B. Minasny, "On digital soil mapping," *Geoderma*, vol. 117, no. 1–2, pp. 3–52, Nov. 2003, doi: 10.1016/S0016-7061(03)00223-4.
[29] L. Yang, Y. Cai, L. Zhang, M. Guo, A. Li, and C. Zhou, "A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102428, Oct. 2021, doi: 10.1016/j.jag.2021.102428.
[30] A. Mora-Vallejo, L. Claessens, J. Stoorvogel, and G. B. M. Heuvelink, "Small scale digital soil mapping in Southeastern Kenya," *CATENA*, vol. 76, no. 1, pp. 44–53, Dec. 2008, doi: 10.1016/j.catena.2008.09.008.
[31] C. Schillaci, S. Saia, and M. Acutis, "Modelling of Soil Organic Carbon in the Mediterranean area: a systematic map," *ROL*, vol. 46, pp. 161–166, Nov. 2018, doi: 10.3301/ROL.2018.68.
[32] M. F. Ghazali, K. Wikantika, A. B. Harto, and A. Kondoh, "Generating soil salinity, soil moisture, soil pH from satellite imagery and its analysis," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 294–306, Jun. 2020, doi: 10.1016/j.inpa.2019.08.003.
[33] A. Mondal, D. Khare, S. Kundu, S. Mondal, S. Mukherjee, and A. Mukhopadhyay, "Spatial soil organic carbon (SOC) prediction by regression kriging using remote sensing data," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 20, no. 1, pp. 61–70, Jun. 2017, doi: 10.1016/j.ejrs.2016.06.004.
[34] M. Ließ, J. Schmidt, and B. Glaser, "Improving the Spatial Prediction of Soil Organic Carbon Stocks in a Complex Tropical Mountain Landscape by Methodological Specifications in Machine Learning Approaches," *PLoS ONE*, vol. 11, no. 4, p. e0153673, Apr. 2016, doi: 10.1371/journal.pone.0153673.
[35] E. Ben-Dor *et al.*, "Using Imaging Spectroscopy to study soil properties," *Remote Sensing of Environment*, vol. 113, pp. S38–S55, Sep. 2009, doi: 10.1016/j.rse.2008.09.019.
[36] E. Vaudour, C. Gomez, Y. Fouad, and P. Lagacherie, "Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems," *Remote Sensing of Environment*, vol. 223, pp. 21–33, Mar. 2019, doi: 10.1016/j.rse.2019.01.006.
[37] K. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine Learning in Agriculture: A Review," *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018, doi: 10.3390/s18082674.
[38] T. T. Nguyen *et al.*, "A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion," *Science of The Total Environment*, vol. 804, p. 150187, Jan. 2022, doi: 10.1016/j.scitotenv.2021.150187.
[39] G. Forkuor, O. K. L. Hounkpatin, G. Welp, and M. Thiel, "High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models," *PLoS ONE*, vol. 12, no. 1, p. e0170478, Jan. 2017, doi: 10.1371/journal.pone.0170478.
[40] A. Bayer, M. Bachmann, A. Müller, and H. Kaufmann, "A Comparison of Feature-Based MLR and PLS Regression Techniques for the Prediction of Three Soil Constituents in a Degraded South African Ecosystem," *Applied and Environmental Soil*

Science, vol. 2012, pp. 1–20, 2012, doi: 10.1155/2012/971252.

[41] S. Zepp, U. Heiden, M. Bachmann, M. Wiesmeier, M. Steininger, and B. van Wesemael, “Estimation of Soil Organic Carbon Contents in Croplands of Bavaria from SCMaP Soil Reflectance Composites,” *Remote Sensing*, vol. 13, no. 16, p. 3141, Aug. 2021, doi: 10.3390/rs13163141.

[42] S. Khanal, J. Fulton, A. Klopfenstein, N. Douridas, and S. Shearer, “Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield,” *Computers and Electronics in Agriculture*, vol. 153, pp. 213–225, Oct. 2018, doi: 10.1016/j.compag.2018.07.016.

[43] R. C. Tryon, “Communality of a variable: Formulation by cluster analysis,” *Psychometrika*, vol. 22, no. 3, pp. 241–260, Sep. 1957, doi: 10.1007/BF02289125.

[44] V. L. Mulder, S. de Bruin, M. E. Schaepman, and T. R. Mayr, “The use of remote sensing in soil and terrain mapping — A review,” *Geoderma*, vol. 162, no. 1–2, pp. 1–19, Apr. 2011, doi: 10.1016/j.geoderma.2010.12.018.

[45] Z. Xu, X. Zhao, X. Guo, and J. Guo, “Deep Learning Application for Predicting Soil Organic Matter Content by VIS-NIR Spectroscopy,” *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–11, Nov. 2019, doi: 10.1155/2019/3563761.

[46] O. Odebiri, O. Mutanga, J. Odindi, K. Peerbhay, and S. Dovey, “Predicting soil organic carbon stocks under commercial forest plantations in KwaZulu-Natal province, South Africa using remotely sensed data,” *GIScience & Remote Sensing*, vol. 57, no. 4, pp. 450–463, May 2020, doi: 10.1080/15481603.2020.1731108.

[47] T. Zhou *et al.*, “Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images,” *Science of The Total Environment*, vol. 755, p. 142661, Feb. 2021, doi: 10.1016/j.scitotenv.2020.142661.

[48] C. Wu, Y. Chen, X. Hong, Z. Liu, and C. Peng, “Evaluating soil nutrients of Dacrydium pectinatum in China using machine learning techniques,” *For. Ecosyst.*, vol. 7, no. 1, p. 30, Dec. 2020, doi: 10.1186/s40663-020-00232-5.

[49] K. Kawamura *et al.*, “Using a One-Dimensional Convolutional Neural Network on Visible and Near-Infrared Spectroscopy to Improve Soil Phosphorus Prediction in Madagascar,” *Remote Sensing*, vol. 13, no. 8, p. 1519, Apr. 2021, doi: 10.3390/rs13081519.

[50] C. Gomez, R. A. Viscarra Rossel, and A. B. McBratney, “Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study,” *Geoderma*, vol. 146, no. 3–4, pp. 403–411, Aug. 2008, doi: 10.1016/j.geoderma.2008.06.011.

[51] R. S. Bricklemyer, R. L. Lawrence, P. R. Miller, and N. Battogtokh, “Monitoring and verifying agricultural practices related to soil carbon sequestration with satellite imagery,” *Agriculture, Ecosystems & Environment*, vol. 118, no. 1–4, pp. 201–210, Jan. 2007, doi: 10.1016/j.agee.2006.05.017.

[52] M. H. Koparan, “Estimating Soil Organic Carbon in Cultivated Soils Using Soil Test Data, Remote Sensing Imagery from Satellites (Landsat 8 and

PlantScope), and Web Soil Survey Data," p. 171.
[53] O. Odebiri, J. Odindi, and O. Mutanga, "Basic and deep learning models in remote sensing of soil organic carbon estimation: A brief review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102389, Oct. 2021, doi: 10.1016/j.jag.2021.102389.
[54] P. D. S. N. Somarathna, B. Minasny, and B. P. Malone, "More Data or a Better Model? Figuring Out What Matters Most for the Spatial Prediction of Soil Carbon," *Soil Science Society of America Journal*, vol. 81, no. 6, pp. 1413–1426, Nov. 2017, doi: 10.2136/sssaj2016.11.0376.
[55] A. M. J.-C. Wadoux, "Using deep learning for multivariate mapping of soil with quantified uncertainty," *Geoderma*, vol. 351, pp. 59–70, Oct. 2019, doi: 10.1016/j.geoderma.2019.05.012.
[56] S. R. Mousavi, F. Sarmadian, A. Rahmani, and S. E. Khamoshi, "DIGITAL SOIL MAPPING WITH REGRESSION TREE CLASSIFICATION APPROACHES BY RS AND GEOMORPHOMETRY COVARIATE IN THE QAZVIN PLAIN, IRAN," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLII-4/W18, pp. 773–777, Oct. 2019, doi: 10.5194/isprs-archives-XLII-4-W18-773-2019.
[57] K. John, I. Abraham Isong, N. Michael Kebonye, E. Okon Ayito, P. Chapman Agyeman, and S. Marcus Afu, "Using Machine Learning Algorithms to Estimate Soil Organic Carbon Variability with Environmental Variables and Soil Nutrient Indicators in an Alluvial Soil," *Land*, vol. 9, no. 12, p. 487, Dec. 2020, doi: 10.3390/land9120487.
[58] W. Zhou *et al.*, "Simulation of Soil Organic Carbon Content Based on Laboratory Spectrum in the Three-Rivers Source Region of China," *Remote Sensing*, vol. 14, no. 6, p. 1521, Mar. 2022, doi: 10.3390/rs14061521.