

Article

Improving the Spatial Prediction of Soil Organic Carbon Content in Two Contrasting Climatic Regions by Stacking Machine Learning Models and Rescanning Covariate Space

Ruhollah Taghizadeh-Mehrjardi ^{1,2}, Karsten Schmidt ^{3,4,*}, Alireza Amirian-Chakan ⁵, Tobias Rentschler ^{1,6}, Mojtaba Zeraatpisheh ⁷, Fereydoon Sarmadian ⁸, Roozbeh Valavi ⁹, Naser Davatgar ¹⁰, Thorsten Behrens ^{1,4,6} and Thomas Scholten ^{1,4,6}

¹ Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, 72070 Tübingen, Germany; ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de (R.T.-M.); t.rentschler@uni-tuebingen.de (T.R.); thorsten.behrens@uni-tuebingen.de (T.B.); thomas.scholten@uni-tuebingen.de (T.S.)

² Faculty of Agriculture and Natural Resources, Ardakan University, Ardakan 8951656767, Iran

³ eScience Center, University of Tübingen, 72070 Tübingen, Germany

⁴ DFG Cluster of Excellence “Machine Learning”, University of Tübingen, 72070 Tübingen, Germany

⁵ Department of Soil Science, Lorestan University, Khorramabad 6815144316, Iran; amirian.ar@lu.ac.ir

⁶ CRC 1070 ResourceCultures, University of Tübingen, 72070 Tübingen, Germany

⁷ Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions, College of Environment and Planning, Henan University, Kaifeng 475004, China; m.zeraatpishe@alumni.iut.ac.ir

⁸ Department of Soil Science, College of Agriculture, University of Tehran, Karaj 77871-31587, Iran; fsarmad@ut.ac.ir

⁹ The Quantitative & Applied Ecology Group, School of BioSciences, The University of Melbourne, Victoria 3010, Australia; rvalavi@student.unimelb.edu.au

¹⁰ Soil & Water Research Institute, Agricultural Research, Education and Extension Organization, Karaj 3177993545, Iran; ndavatgar@swri.ir

* Correspondence: karsten.schmidt@uni-tuebingen.de

Received: 21 February 2020; Accepted: 26 March 2020; Published: 29 March 2020



Abstract: Understanding the spatial distribution of soil organic carbon (SOC) content over different climatic regions will enhance our knowledge of carbon gains and losses due to climatic change. However, little is known about the SOC content in the contrasting arid and sub-humid regions of Iran, whose complex SOC–landscape relationships pose a challenge to spatial analysis. Machine learning (ML) models with a digital soil mapping framework can solve such complex relationships. Current research focusses on ensemble ML models to increase the accuracy of prediction. The usual ensemble method is boosting or weighted averaging. This study proposes a novel ensemble technique: the stacking of multiple ML models through a meta-learning model. In addition, we tested the ensemble through rescanning the covariate space to maximize the prediction accuracy. We first applied six state-of-the-art ML models (i.e., Cubist, random forests (RF), extreme gradient boosting (XGBoost), classical artificial neural network models (ANN), neural network ensemble based on model averaging (AvNNNet), and deep learning neural networks (DNN)) to predict and map the spatial distribution of SOC content at six soil depth intervals for both regions. In addition, the stacking of multiple ML models through a meta-learning model with/without rescanning the covariate space were tested and applied to maximize the prediction accuracy. Out of six ML models, the DNN resulted in the best modeling accuracies, followed by RF, XGBoost, AvNNNet, ANN, and Cubist. Importantly, the stacking of models indicated a significant improvement in the prediction of SOC content, especially when combined with rescanning the covariate space. For instance, the RMSE values for SOC content prediction of the upper 0–5 cm of the soil profiles of the arid site and the sub-humid site by the

proposed stacking approaches were 17% and 9% respectively, less than that obtained by the DNN models—the best individual model. This indicates that rescanning the original covariate space by a meta-learning model can extract more information and improve the SOC content prediction accuracy. Overall, our results suggest that the stacking of diverse sets of models could be used to more accurately estimate the spatial distribution of SOC content in different climatic regions.

Keywords: digital soil mapping; machine learning models; stacking of models; spatial block cross-validation; deep learning

1. Introduction

Soil organic carbon (SOC) storage is a key function of soils, influencing soil physicochemical properties [1,2], e.g., soil water storage capacity, nutrient holding capacity, and infiltration rate. As the world's soils contain more organic carbon than the atmosphere and the biosphere together, soils are considered to be a crucial pool in the global carbon cycle [3]. Thus, accurate information on the spatial distribution of SOC is vital to estimate and predict greenhouse gas emissions and physicochemical functions of soils [4,5]. Such information is most important in arid and semi-arid areas where soils tend to have low levels of organic carbon [6,7] compared to the humid region. These sensitive and fragile ecosystems are less resilient against climate change and, therefore, more vulnerable to desertification [8,9].

Legacy soil maps based on traditional soil mapping approaches are the most common sources for acquiring data and information on soils in Iran [9]. The qualitative nature and coarse scales of the available maps make these maps impractical for quantitative studies and a detailed understanding of the spatial variations of soil properties [10–12]. Furthermore, traditional soil mapping approaches are time-consuming and expensive [13]. Digital soil mapping approaches based on the scorpan concept [14] have become a standard approach to generate new soil data to overcome the limitations arising from the legacy soil maps. Digital soil mapping provides a quantitative-empirical framework for predicting soil properties and classes from spatially referenced covariates using appropriate machine learning (ML) models [5].

Several ML models have successfully linked SOC to environmental covariates to extrapolate SOC to unknown locations [15–29]. Some of the most popular models are multivariate regression, classical artificial neural networks [13], support vector regression [20], regression trees [17,20], and random forests [15,20,30]. Recently, deep neural networks based on deep learning approaches were used to solve highly complex soil–landscape problems [31–34]. Padarian et al. [33] and Wadoux et al. [34] predicted and mapped SOC in Chile and France respectively, using deep learning methods.

One commonly applied technique to improve the predictive capacity and to decrease the variance of the individual ML model is the ensemble model—bagging, boosting, and stacking approach [35]. Bagging is a simple and very powerful ensemble method. It generates m new training sets and then m models are fitted to the datasets. Their prediction results are combined by averaging the output or voting. Boosting refers to a group of algorithms that use weighted averages to turn weak learners into stronger learners. Other ensemble techniques include model averaging [36,37].

The stacking approaches combine different types of models (lower level) through a meta-learning model (higher level) to maximize the generalization accuracy [38]. Unlike bagging, boosting, and averaging methods, stacking ensemble modeling is rarely explored in digital soil mapping. Nevertheless, stacking often performs better than all individual models, especially when combined with rescanning the original covariate space [39]. For instance, Tajik et al. [40], Zhou et al. [41], and Chen et al. [42] recently evaluated the efficacy of the ensemble models—by averaging the model predictions—to predict the spatial variation of soil properties in Iran, China, and France, respectively.

Although different ML models were implemented in order to predict and map SOC [5], their performances are inconsistent in various SOC studies. To the best of our knowledge, there is no study to conduct digital mapping of SOC content using stacking approaches in different climatic conditions. Thus, the authors suggest combining the ML models with the rescanning of the original covariate space to explore if it works better than the standard stacking of individual models. Furthermore, so far, only a few studies have used deep learning models for digital soil mapping, with the notable exceptions of References [31,33,34], and a comparison with other models is still needed. This study therefore fills a void related to digital soil mapping applications of deep learning methods. Finally, there is a lack of understanding concerning the prediction of SOC content under different climatic regimes in Iran, which has a vast territory and diverse climates. Most studies conducted on SOC content in Iran only consider a single climatic influence [7,22,30].

Hence, the main objective of this study is to evaluate and compare stacking ensemble approaches with six ML models in order to predict and map the spatial distribution of SOC content for two areas with contrasting climate (i.e., arid and sub-humid) in Iran. The models include Cubist, random forests (RF), extreme gradient boosting (XGBoost), classical artificial neural network models (ANN), neural network ensemble based on model averaging (AvNNet), as well as deep learning neural networks (DNN). We further tried to identify the controlling factors of the spatial distribution patterns of SOC content in the contrasting climatic conditions, which has rarely been reported in Iran.

2. Materials and Methods

2.1. Study Sites

This study was conducted at two sites located in central (Yazd province) and northern (Gilan province) Iran (Figure 1a). The study sites comprise two diverse climatic regions (Figure 1b), which are arid in Yazd province and sub-humid in Gilan province [43]. The general climate conditions of selected sites are presented in Table 1.

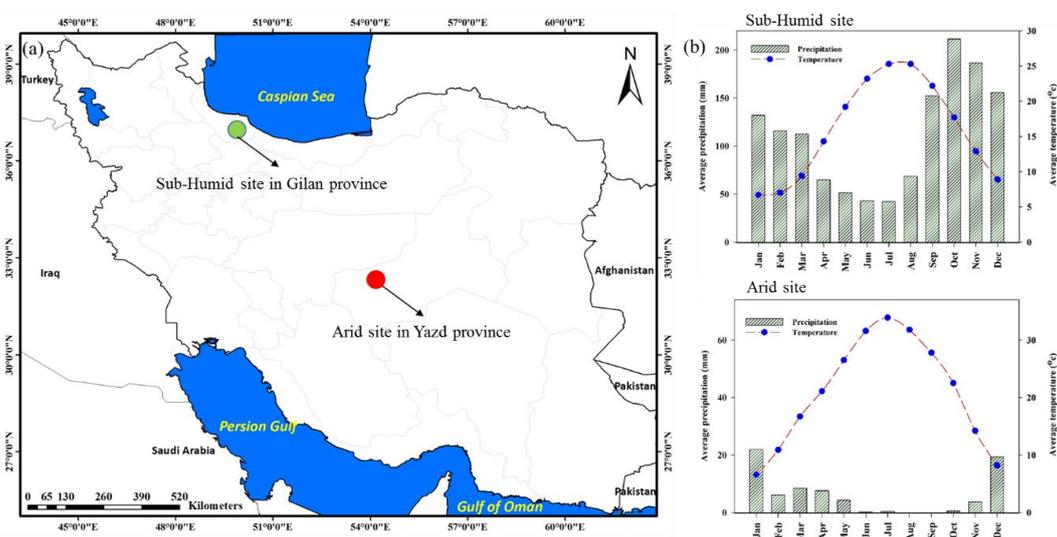


Figure 1. Study areas in Iran (a), and the ombrothermic graphs of the arid and sub-humid sites (b).

Table 1. Study sites and data collection details.

Site Names	Area (km ²)	Soil Types	Climate Conditions	Precipitation (mm/year)	Elevation (m)	Samples (no.)
Arid site	720	Solonchaks, Gypsisols and Regosols	Arid	75	944–1944	154
Sub-Humid site	3000	Kastanozems, Cambisols and Chernozems	Sub-Humid	1200	–26–700	99

The arid site is located in the Yazd province in central Iran and covers 720 km². The average annual precipitation, temperature, and annual potential evaporation are 75 mm, 18.5 °C, and 3483 mm, respectively. The soil moisture and temperature regimes are aridic and thermic [43,44]. The elevation ranges from 944 to 1944 m above sea level. The main land use types consist of cropland (pistachio nuts and wheat) and grassland. The major physiographic units from East to West are alluvial fans, coalescing alluvial fans (bajadas), salt plains, and gypsiferous hills. The predominant soils in the study area [43,44] are Solonchaks with ~40%, Gypsisols with ~40%, and Regosols with ~20% of the area [43,44].

The sub-humid site is located in the Gilan province in northern Iran and covers 3000 km². The climate is sub-humid, and the average annual precipitation, temperature, and annual potential evaporation are 1200 mm, 15.6 °C, and 796 mm, respectively. The soil moisture and temperature regimes are udic and thermic [43]. The elevation ranges from −26 to 700 m above sea level. The main land use types consist of cropland (rice) and forest (oak, beech, and elm). Except for the southern parts of the study area, where piedmonts and hills dominate, the topography of the area is mainly flat. Predominant soils of the study area [43] are Kastanozems with ~70%, Cambisols with ~25%, and Chernozems with ~5% of the area.

2.2. Data Collection and Soil Sample Analysis

For the purpose of digital soil mapping, a well-distributed sample set is needed. We used the conditioned Latin Hypercube Sampling, which provides an optimal stratification of the covariate space [45,46], to select representative sample locations based on the covariates [47–52]. We selected a total of 154 and 99 soil profiles for the arid and sub-humid sites, respectively. Soil samples were collected from the genetic horizons of each profile down to a depth of 2 m. Air-dried soil samples were ground and sieved (<0.5 mm), and the SOC content (%) was determined by wet oxidation [53].

Sampling by genetic horizons means that samples do not come from consistent depth intervals in all locations. Therefore, we used an equal-area spline function [54] to harmonize SOC data and estimate the vertical variation of SOC content. The equal-area spline function was fitted to each profile. Then, the values of SOC content were obtained by the integration of the splines to the defined depth intervals. We estimated the SOC at six depth intervals of 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm, in accordance with the standard depths specified by the GlobalSoilMap project [55].

2.3. Covariates Used for the Development of ML Models

We used a set of 28 covariates (Table 2) as predictors [5,14] representing potential environmental drivers of the spatial and vertical distribution of SOC content. Based on the understanding of the factors affecting the SOC content distribution in the two study areas [7,22,50] and literature reviews [5,56], the covariates were obtained and derived from a digital elevation model (DEM) and remotely sensed satellite data.

The Shuttle Radar Topography Mission (SRTM) DEM with a resolution of 30 × 30 m was used for the terrain analysis [57]. The DEM was preprocessed to fill the sinks and pits before ten terrain attributes were calculated (Table 2) using SAGA GIS (System for Automated Geoscientific Analyses) [58]. These are elevation, wetness index, catchment area, catchment slope, multi-resolution valley bottom flatness index, valley depth, plane curvature, profile curvature, general curvature, and total insolation.

The remote sensing (RS)-based covariates were derived and calculated based on the median values of 127 cloud-free Landsat-8 [59] and Sentinel-2 [60] images taken during 2016 under clear and dry weather conditions during the spring/summer season using the Google Earth Engine environment [61]. In general, we used six spectral bands of Landsat-8 (B2, B3, B4, B5, B6, and B7) and ten spectral bands of Sentinel-2 (B2, B3, B4, B5, B6, B7, B8, B8a, B11, and B12), respectively [47,48]. Additionally, we calculated the normalized difference vegetation index (NDVI) using spectral bands of both Landsat-8 and Sentinel-2 [62–64].

All covariates were rescaled using Z-score standardization and resampled in order to have similar scale and the same cell size of 30×30 m.

Table 2. Covariates used for the development of machine learning (ML) models.

No.	Definition	Abbreviation
Terrain-based covariates		
1	Elevation	Elev
2	Wetness Index	WI
3	Catchments area	Ca.Area
4	Catchment Slope	Ca.Slop
5	Multi-resolution Valley Bottom Flatness Index	MrVBF
6	Valley Depth	Vally.D
7	Plane Curvature	Pl.Cur
8	Profile Curvature	Pr.Cur
9	General Curvature	Ge.Cur
10	Total Insolation	To.In
Remote Sensing-based covariates		
11	Blue band of Landsat-8 (0.482 μm)	B2.L
12	Green band of Landsat-8 (0.561 μm)	B3.L
13	Red band of Landsat-8 (0.654 μm)	B4.L
14	Near infrared band of Landsat-8 (0.864 μm)	B5.L
15	Shortwave Infrared-1 band of Landsat-8 (1.608 μm)	B6.L
16	Shortwave Infrared-2 band of Landsat-8 (2.200 μm)	B7.L
17	Blue band of Sentinel-2 (0.490 μm)	B2.S
18	Green band of Sentinel-2 (0.560 μm)	B3.S
19	Red band of Sentinel-2 (0.665 μm)	B4.S
20	Vegetation Red Edge of Sentinel-2 (0.705 μm)	B5.S
21	Vegetation Red Edge of Sentinel-2 (0.740 μm)	B6.S
22	Vegetation Red Edge of Sentinel-2 (0.783 μm)	B7.S
23	Near infrared band of Sentinel-2 (0.842 μm)	B8.S
24	Vegetation Red Edge of Sentinel-2 (0.865 μm)	B8a.S
25	Shortwave IR-1 band of Sentinel-2 (1.610 μm)	B11.S
26	Shortwave IR-2 band of Sentinel-2 (2.190 μm)	B12.S
27	Normalized difference vegetation index (Landsat-8 based)	NDVI.L
28	Normalized difference vegetation index (Sentinel-2 based)	NDVI.S

2.4. Covariate Selection

In this study, the Boruta algorithm [65] was implemented with the random forests (RF) classifier in the R statistical package [66] to rank the most important covariates for predicting SOC content at six depths. The algorithm consists of the following steps:

- i. The covariate space is extended by adding randomly permuted existing covariates (pC) in order to remove their correlation with SOC content,
- ii. A RF prediction using the extended covariate space (i.e., covariates and permuted covariates) is performed to predict SOC content at six standard depths,
- iii. The Z-score, which is an indicator of the importance of all covariates, is computed,
- iv. The maximum Z-score (MZSA) among the pC's is defined,
- v. A hit is assigned to all covariates that scored better than MZSA,
- vi. A two-test of equality is performed for undetermined important covariates,
- vii. The original covariates are respectively flagged as "unimportant" or "important" if they have significant lower or higher scores than MZSA,
- viii. All permuted covariates are removed,
- ix. Repeating the procedure.

In this study, based on Z-score values [67], we grouped the ability of covariates to explain SOC content variability into 4 classes: weakly relevant ($Z < 5$), slightly relevant ($5 < Z < 10$), moderately relevant ($10 < Z < 15$), and relevant ($Z > 15$).

2.5. Stacked Generalization

Stacked generalization or simply stacking is an ensemble approach that combines the outcomes of different ML models in a single model to maximize the generalization accuracy [35,37]. Usually, as illustrated in Figure 2, there are two levels in a stacking framework: level 0 and level 1, consisting of several base models and one meta-learning model. Meta-learning models in level 1 use the prediction of the response variables that are estimated by several base models in level 0 in order to generate a final prediction. In other words, the model in level 1 learns with the predictions of the models of level 0.

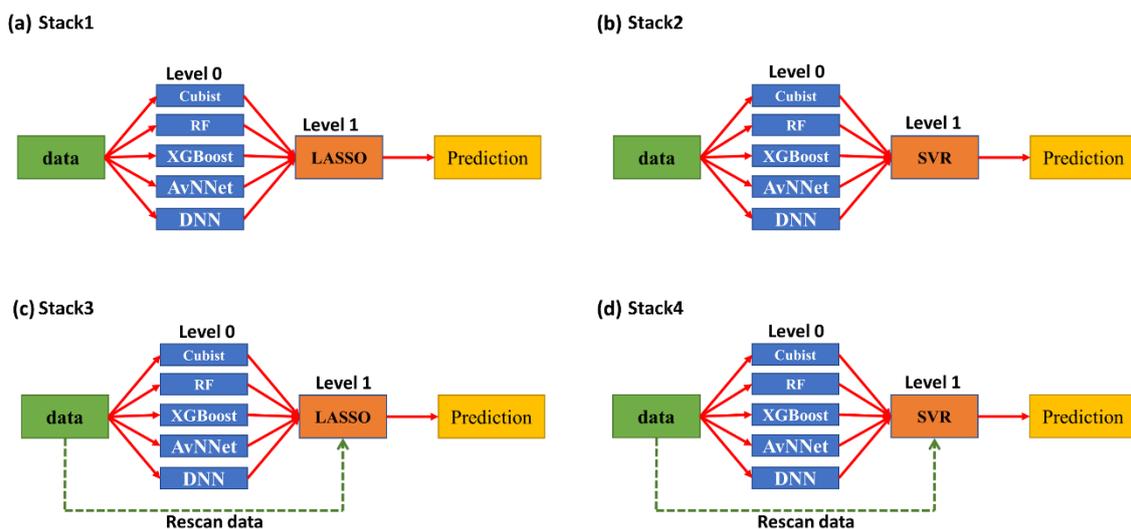


Figure 2. General framework of stacking approaches used in this study. (a) Cubist/Random Forests (RF)/extreme gradient boosting (XGBoost)/classical artificial neural network models (ANN)/neural network ensemble based on model averaging (AvNNet)/deep learning neural networks (DNN) + least absolute shrinkage and selection operator (LASSO); Stack1; (b) Cubist/RF/XGBoost/ANN/AvNNet/DNN + support vector regression (SVR); Stack2; (c) Cubist/RF/XGBoost/ANN/AvNNet/DNN + LASSO + rescan mode; Stack3; (d) Cubist/RF/XGBoost/ANN/AvNNet/DNN + SVR + rescan mode; Stack4.

In this study, we used six ML models (Cubist, RF, XGBoost, ANN, AvNNet, and DNN) in level 0. Conventionally, the meta-learning model in level 1 is based on a weighted average method or a linear regression model [36,37]. In this study, we applied two new meta-learning models (least absolute shrinkage and selection operator (LASSO) and support vector regression (SVR)) in level 1.

Furthermore, we introduced two modes of stacking: the standard mode (Stack1 and Stack2) and the rescan mode (Stack3 and Stack4), as shown in Figure 2 [68]. Special attention should be given to the fact that in the rescan mode, we allow the model in level 1 (LASSO and SVR) to learn again from the original covariate space in order to extract some missing information. Practically, in the standard mode, we used the predicted SOC contents of individual models (Level 0: such as RF and ANN) as the predictor variables for meta-learning models (Level 1: such as LASSO and SVR). In the rescan mode, we used both the primary covariates (such as NDVI and MrVBF) and predicted SOC contents of individual models (Level 0: such as RF and ANN) as the predictor variables for meta-learning models (Level 1: such as LASSO and SVR). In summary, we tested four stacking approaches. A more detailed account of models used in level 0 and level 1 is given in the following sections.

2.5.1. The Individual ML Models in Level 0

Cubist [69] is an extension of the M5 algorithm. It is similar to common regression trees, but the terminal nodes contain linear least square models of covariates used in the previous intermediate node [70] rather than discrete values. Also, there are intermediate linear least squares models at each step of the tree, which are used to adjust the final prediction. Cubist uses “if-then” rules to partition

the training data [71]. Whenever the conditions of a rule are satisfied, the associated linear least square model is used to predict the response [72].

RF [73] is an ensemble technique based on the well-known classification and regression tree approach (CART). The ensemble is generated by averaging several trees based on different bootstrap sample sets selected from the training data. Further, only a random subset of covariates is evaluated at each node. RF with a large number of trees is robust against overfitting, noise, as well as non-informative and correlated features. RF has been used in various DSM studies over the past decade [74–76] and for many other environmental problems [77]. Extreme gradient boosting (XGBoost) [78] is also a tree-based ensemble method. However, instead of independent trees and averaging the individual predictions, the XGBoost creates a number of decision trees sequentially. The trees are generated by using the residuals or prediction errors of the previous tree model, thus the algorithm focuses more on samples with higher uncertainty. Finally, all generated models are added together to calculate the outcome [79].

The most common ANNs, also known as multi-layer perceptron (MLP), consist of three layers, i.e., an input layer, a hidden layer, and an output layer. Each hidden unit combines all input units of the input layer, where all connections are associated with a weight. Further, an activation function is applied to the sum of weighted unit inputs. The output layer is calculated the same way as the hidden units, but with input from the hidden units. For the MLP with one hidden layer, we used the sigmoid function as the activation function in nnet package [80,81]. The network was trained through back-propagation using the Levenberg–Marquardt algorithm with 150 iterations [82].

AvNNet is similar to MLP, but multiple neural network models with the same topology are used to predict the response. The models can be different either due to different random number seeds to initialize the network or by fitting the models on bootstrap samples of the original training set (i.e., bagging the neural network). All the resulting models are used for prediction [81]. For regression, the outputs from each network are averaged [83]. The idea behind AvNNet is that we usually train different ANN models for the same problem in order to figure out the best model that produces the best validation statistics. However, instead of choosing the best model, it is possible to combine all models in order to improve the generalization power of a single neural network [84]. In this study, we used the AvNNet model [80,81] in the Caret package [85]. We note that the tuning parameters used for MLP were kept the same for AvNNet.

Deep learning neural networks (DNN) use the MLP structure, but have more hidden layers and a more hierarchical structure [86]. DNNs with multiple hidden layers, as shown in Appendix A (Figure A1), have a huge number of hyper-parameters (e.g., optimization algorithm, learning rate, network weight initialization, hidden layers activation function, output activation function, L2 regularization, dropout regularization, and the number of nodes in the hidden layers) [87]. The hyper-parameters potentially allow DNNs to perform better in solving the complex problems compared to the other ML models [88]. Sometimes, however, a lack of control over the learning process of the DNNs may lead to overfitting [32]. One approach, which is also used in this study to avoid or reduce overfitting, is to use a technique called Dropout [89]. Dropout randomly mutes neurons of the hidden layers. This dropout is applied to each of the n training steps, resulting in n different networks that are finally averaged for prediction [90]. For predictions, the ensemble of sparse networks resulting from the dropout process is averaged using the geometric mean of the input weights of the neurons. In this study, for DNNs we used the H2O package [91] with the rectifier function as a non-linear transformation and the Stochastic Gradient Descent (SGD) as the optimization algorithm. Furthermore, in order to save training time, an early stopping was used if no changes in the loss were observed after 150 epochs. Appendix A (Tables A1 and A2) shows the specifications used for DNN in this study, namely: hidden layers, size, network weight initialization, learning rate, and dropout regularization.

2.5.2. Meta-Learning Models in Level 1

The least absolute shrinkage and selection operator (LASSO) is a regularized linear model. It adds a regularization term as a cost function to a linear model, to reduce its degrees of freedom. To achieve

this, the lasso regression performs feature selection by eliminating the weights of the least important predictors. For the Lasso modeling, we used the glmnet package [92].

Support vector machines are a kernel method for classification [93] and regression problems [94]. The input data is transformed into a high-dimensional feature space with a predefined kernel function. In the high-dimensional feature space, a linear regression hyperplane is derived for non-linear relationships. Then, the hyperplane is back-transformed to non-linear space. The kernel used in this study is a radial basis function. The e1071 package [95] was used for radial SVR modeling.

2.6. Optimizing the Hyper-Parameters of Machine Learning Models

We applied a grid-learning method to estimate the best model-parameter by testing different ranges of the model parameters listed in Table 3. Importantly, these hyper-parameters are the most likely parameters to have the largest effect on the performance of the ML models. All other hyper-parameters were set to their defaults [96]. Based on the most relevant parameters, we tuned each model individually and evaluated the prediction performance. Additionally, we combined the grid-learning method with a spatial block cross-validation strategy with the aim to reduce the spatial autocorrelation effect of close neighbors and to choose the optimal model parameter. In this study, we constructed 10 folds for our block cross-validation using R package blockCV [97], in which several spatial blocks can be assigned to a fold (Figure 3). The block-to-fold assignment in this package was done by a repeated random approach that tries to find the most evenly distributed number of observations in each fold. Thus, the observations are separated spatially and in each fold as close as possible to the typical 10-fold cross-validation approach.

Table 3. Hyper-parameters of ML models tuned in this study.

ML Models	Hyper-Parameters	Definition	Defined Parameters
Cubist	committees	the number of model trees	1–100
	neighbors	the number of nearest neighbors	0–9
	booster	the type of model	gbtree
	max_depth	the depth of tree	3–10
XGboost	min_child_weight	the minimum sum of weights of all observations	0–5
	colsample_bytree	the number of variables supplied to a tree	0.5–1
	subsample	the number of samples supplied to a tree	0.5–1
	eta	learning rate	0.01–0.5
RF	Mtry	the number of input variables	1–30
	Ntree	the number of trees	100–3000
ANN	decay	learning rate	0.001–0.05
	size	the number of neurons in hidden layer	1–10
AvNNet	Repeats	the number of MLP with different random number seeds	3–20
	hidden size	the number of hidden layers	2–10
DNN	network weight initialization	the initialized weight of networks	uniform/he_normal
	learning rate	that controls adjusting the weights of the network	0.001–0.05
	dropout regularization	the amount of the neurons that are randomly dropped	0.2–0.8
SVM	Kernel type	the kernel function	RBF
	C	the penalty parameter	0.01–100
Lasso	σ	the bandwidth parameter	0.01–100
	lambda	the shrinkage parameter	1–150

MLP: multilayer perceptron; RBF: radian basis function.

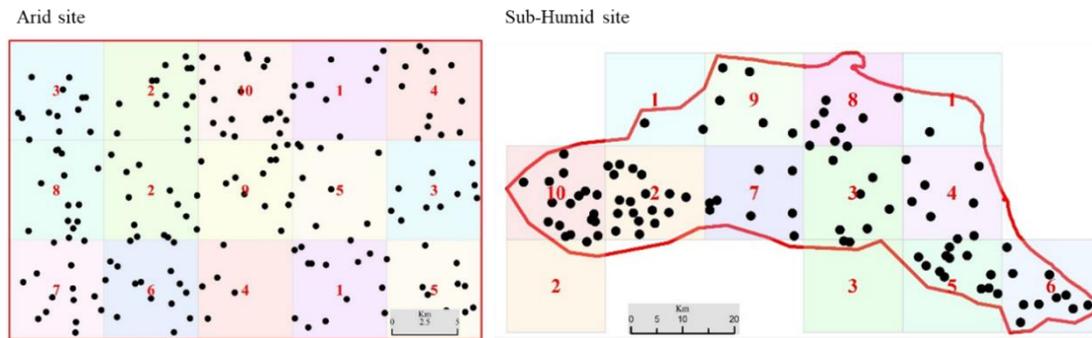


Figure 3. Illustration of the spatial blocking strategy in two regions. The numbers in the blocks are fold numbers, showing allocation of blocks to folds.

2.7. Statistical Evaluation

In this study, four common performance metrics [98], namely root mean squared error (RMSE), normalized root mean squared error (nRMSE), coefficient of determination (R^2), and Ratio of Performance to InterQuartile distance (RPIQ) were used. RMSE indicates the accuracy of the model prediction. nRMSE is without unit and the standardized form of RMSE and well suited for inter-model comparisons. The coefficient of determination (R^2) varies between 0 and 1 and indicates the closeness of the observed values to the fitted regression line or the proportion of variance explained by the independent predictors. RPIQ compares the interquartile range to the RMSE [99]. The greater the RPIQ indicates the better the model's predictive capacity.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - P_o)^2} \quad (1)$$

$$nRMSE = \frac{RMSE}{\bar{O}} \quad (2)$$

$$R^2 = \left(\frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (P_i - \bar{P})^2}} \right)^2 \quad (3)$$

$$RPIQ = \frac{Q_3 - Q_1}{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - P_o)^2}} \quad (4)$$

where P_i and O_i are the predicted and observed SOC values at the i th location, n is the number of data points, \bar{P} and \bar{O} denote the means for the predicted and observed SOC, and Q_1 and Q_3 are the first and third quartiles, respectively.

3. Results and Discussion

3.1. Summary Statistics of SOC Content

The descriptive statistics of the SOC content at six depth intervals across the two study areas are presented in Table 4. For the arid site, the mean SOC content varied from 0.18% to 0.33%, whereas in the sub-humid site, it ranged from 1.46% to 4.09% (Table 4). The lower and upper limits of the mean at 95% varied from 0.16 to 0.39 for the arid site, whereas in the sub-humid site, it ranged from 1.24% to 4.38%. This indicates a high variability of SOC content across the two sites. The highest variability in SOC content was found at the arid site with a coefficient of variation from 60.39% for the 60 to 100 cm depth to 128.59% for the first depth increment (0–5 cm). Similarly, the sub-humid site showed a high variability of SOC content with a coefficient of variation from 37.15% for the 0 to 5 cm depth to 78.66%

for the deepest depth increment (100–200 cm). The arid site, in contrast to the sub-humid site, tended to have higher variability in SOC content at the upmost depth increments [6,100].

Table 4. Descriptive statistics of soil organic carbon (SOC) content at six standard depths in two regions.

Soil Depth	SOC (%)						
	Min	Max	Mean	Lower	Upper	SD	CV
Arid site							
0–5 cm	0.03	2.34	0.33	0.26	0.39	0.42	128.59
5–15 cm	0.04	2.21	0.31	0.25	0.37	0.39	124.56
15–30 cm	0.06	1.69	0.27	0.23	0.32	0.30	110.24
30–60 cm	0.02	1.11	0.21	0.19	0.24	0.17	77.28
60–100 cm	0.01	0.75	0.18	0.16	0.19	0.11	60.39
100–200 cm	0.01	1.00	0.18	0.16	0.20	0.14	78.20
Sub-Humid site							
0–5 cm	1.36	9.93	4.09	3.79	4.38	1.52	37.15
5–15 cm	1.28	9.51	3.68	3.41	3.95	1.39	37.89
15–30 cm	0.68	8.01	2.59	2.34	2.85	1.30	50.27
30–60 cm	0.41	5.65	1.55	1.35	1.75	1.03	66.26
60–100 cm	0.07	5.65	1.46	1.24	1.69	1.15	78.21
100–200 cm	0.07	5.65	1.47	1.24	1.69	1.15	78.66

Min: minimum; Max: maximum; SD: standard deviation; CV: coefficient of variation; Lower and Upper: the lower and upper limits of the mean at 95%.

The results of the mean SOC content comparisons for arid and sub-humid sites are shown in Figure 4. At the sub-humid site, the upper three layers (0–30 cm) are significantly different in terms of mean SOC content, whereas the lower depth intervals show no significant differences in SOC content. This result indicates more variation in the vertical distribution of SOC content at the topsoil compared to the subsoil. The mean SOC contents in the arid site show a relatively different trend compared to the sub-humid site. At the arid site, there are no significant differences between the three upper horizons. This indicated that for the top three layers (0–30 cm), depths intervals had no significant effect on SOC content.

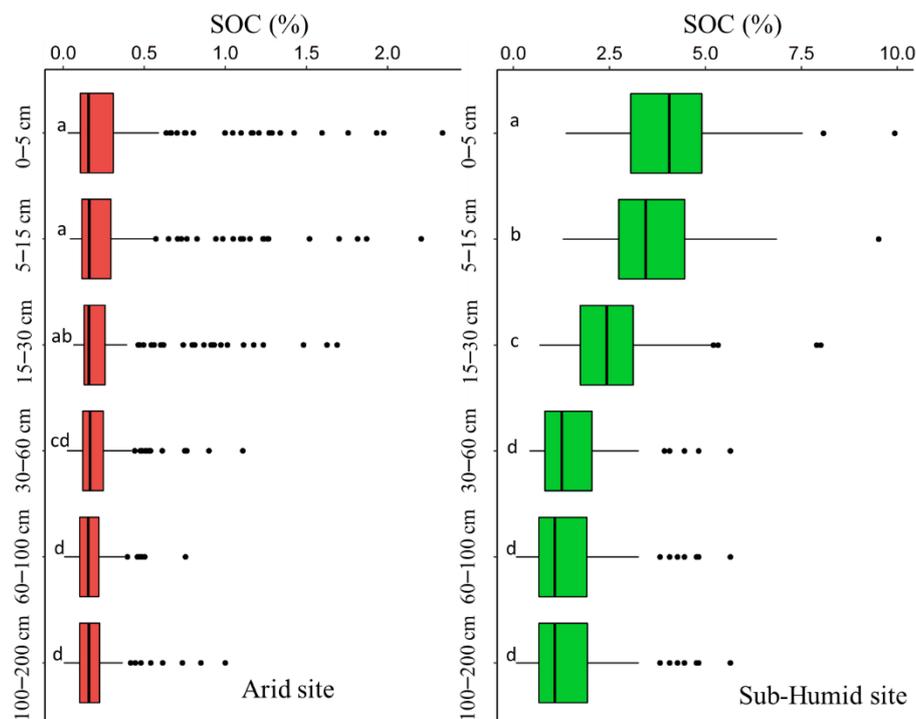


Figure 4. Box plot of SOC content at six standard depths at the two regions. Different letters (a:d) indicate a significant difference of mean SOC content at the 0.05 level.

A decreasing trend in SOC content with increasing depth was found on both sites. This is much more evident at the sub-humid site (Figure 4). The SOC content in both arid and sub-humid areas at the surface layer (0–5 cm) were about 1.8 and 2.8 times higher than the SOC content in the depth of 100 to 200 cm (Table 4). Several studies reported that SOC content in the topsoil was more abundant than in the subsoil [7,22,51,54,101].

3.2. Importance of Covariates

The selected covariates for prediction of the SOC content at the two sites at all specific depths are presented in Figure 5. The numbers indicate Z-scores and the intensity of colors from light to dark represents the values of Z-scores from low to high, respectively. The covariates used to predict SOC content showed a varying level of importance in the models. Results indicated that the covariates in the arid site were weakly to moderately relevant to SOC content. The Z-score varied from 0.40 to 10.60 for valley depth (SOC 100–200 cm) and the near infrared band of Sentinel-2 (SOC 5–15 cm), respectively. In the sub-humid site, it varied from 0.20 to 23.80 for the variables of total insolation (SOC 100–200 cm) and the Green band of Sentinel-2 (SOC 0–5 cm), respectively.

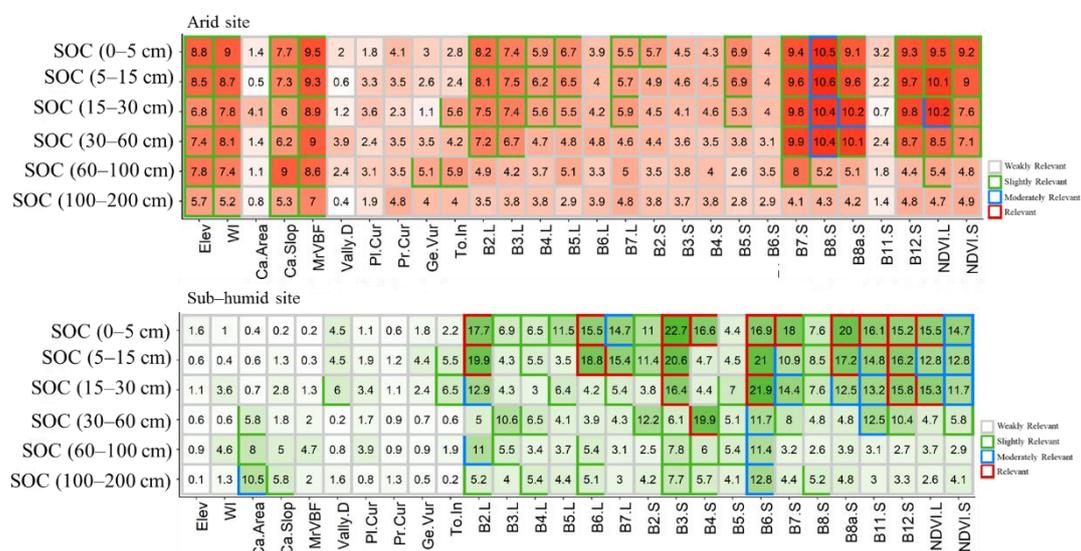


Figure 5. Optimal covariate selection using Boruta algorithm for SOC content at six standard depths at the two regions. The covariate is defined in Table 2.

For the arid site, Figure 5 shows that all covariates were moderately relevant (e.g., NIR band of Sentinel-2) or slightly relevant (e.g., Elevation and wetness index) to predict SOC content, at least at one soil depth interval, except for nine covariates (e.g., Catchments area). Four covariates (e.g., MrVBF) were identified as slightly relevant at all six soil depths (Figure 5). In the sub-humid site, however, only one covariate (vegetation red edge of Sentinel-2) was important at all six depth intervals and six covariates (e.g., Elevation) out of 28 were weakly relevant in predicting SOC at any depth interval (Figure 5). It should be highlighted that all other covariates were classified as slightly relevant (e.g., Catchment slope), moderately relevant (e.g., NDVI), and relevant (e.g., Green band of Sentinel-2) variables in predicting SOC content at all six depth intervals.

For the arid site, according to Figure 5, one can further conclude that both terrain- and RS-based covariates play an important role in the prediction of SOC content at six depth intervals, which is in line with the study of Wang et al. [19,102]. Importantly, the RS-based covariates (e.g., NIR band of Sentinel-2) were more important in predicting organic carbon of surface soils in comparison with the terrain-based covariates (e.g., MrVBF). This can be because most of the area was bare except for the center of the area and, thus, the remotely sensed data could represent the spectral behavior of soil at the surface [21,22]. However, terrain-based covariates were more relevant for the SOC content

prediction at the depth of 60 to 200 cm, compared to the RS-based covariates. This is particularly true for elevation (Elev), wetness index (WI), catchment slope (Ca.Slop), and multi-resolution valley bottom flatness index (MrVBF). This local topography could define the rate of soil erosion and deposition and also the amount of incoming solar radiation, which can determine SOC distribution. In addition, the influence of terrain-based covariates on predicting SOC content can be related to the other soil properties' variations as well as the existence of cultivation in the lower elevated areas (e.g., center of the area). The major factor that controls the input, decomposition, and stabilization of organic matter into the soil is the cultivation in the center of the area. The cultivation explained the amount and variation of SOC content in the study area [76]. Consequently, in the arid site, the SOC distribution may be partly characterized by topography. In line with our results, several studies [7,22,54] reported the importance of terrain-based covariates such as WI, MrVBF, slope, and elevation for the prediction of SOC at different soil depths, because they reflect important geomorphological units, e.g., deeper soil development at the valley bottom with steep slopes surrounding as a result of erosion processes.

Whereas, for the sub-humid site, the terrain-based covariates are not controlling factors on SOC content variability, with a notable exception of catchments area in the lower depth intervals (30–200 cm). These results were to be expected because the study area in the northern parts of Iran has almost flat terrain [83]. However, the finding revealed the importance of RS-based covariates on SOC content variability in the sub-humid site (Figure 5). For instance, we found NDVI as an important predictor for SOC content at surface layers of soils [47,48]. The remote-sensed vegetation parameters and NDVI are commonly considered as good indicators of primary and ecological productivity. Hence, the remote-sensed data have been effectively applied to predict SOC contents [48]. Consequently, in the sub-humid site, these results show that the vegetation has the most effective influence on topsoil SOC. Furthermore, spectral bands of Sentinel-2 had a substantial influence on the estimation of SOC contents at the 0 to 30 cm soil depth in the sub-humid site, as shown in Figure 5. This shows the high potential of spectral data in detecting SOC and its variation. Several studies [47,103] also reported a great contribution of Sentinel-2 images to predict SOC contents of soils in the Czech Republic and France. For instance, Gholizadeh et al. [48] proved the potential Sentinel-2 image to capture the variation of SOC, especially where SOC levels were relatively high. They reported that the best SOC and Sentinel-2 spectral bands' correlations were obtained from B4 and B5 followed by B11 and B12.

General speaking, RS-based covariates mostly represent surface features indicating land use/cover, parent materials, and any other features influencing spectral behaviors. Terrain-based covariates mostly represent relief parameters which contribute to erosional and depositional process as well as to water and fine particles' accumulation and redistribution. As discussed above, both terrain- and RS-based covariates can potentially explain the variation of SOC content at the two study sites (Figure 5). Nevertheless, the relative influence of covariates is distinct at the arid and sub-humid regions. According to Figure 5, our results revealed that RS-based covariates could better explain the variation of SOC content in the sub-humid site compared to the arid site. This is expected because vegetation cover by affecting land reflectance in visible and infrared [21] makes RS-based covariates promising explanatory variables to explain SOC contents and variations in the sub-humid regions [22], because a simple correlation between the photosynthetic activity of plants and a higher amount of source material (RS-based proxies) can lead to a higher accumulation of organic carbon. In other words, the effect of vegetation on SOC content might be represented by RS-based covariates. However, terrain-based covariates by controlling erosional and depositional processes were more successful in the arid site compared to the sub-humid site to explain SOC content variations. This could be due to the fact that the climate is uniform within the arid site and therefore is not a controlling factor. However, topography is a significant factor to explain SOC content variations. The least importance to terrain-based covariates in sub-humid sites is expected and attributed to the fact that the flatness in the sub-humid site minimizes the effect of topography and elevation on the soils [104]. In addition, the relative effect of covariates varies with depth, indicating that the mechanisms involved in SOC

stabilization and dynamics at different depths may be various. Basically, the cumulative effect of these variables influences the SOC contents.

3.3. Performances of the Individual ML Models

For both sites, the performance of six individual ML models used in level 0 in terms of R^2 , RMSE, and RPIQ at six depth intervals was as follows: DNN > RF > XGBoost > AvNNNet > ANN > Cubist (Tables 5 and 6). Our results indicated that the RF models can well predict SOC content in the two study sites. In agreement with our findings, Keskin et al. [67] reported that RF resulted in the lowest RMSE for SOC prediction compared to the other ML models because of random selection of variables during tree building and assembly. Our experience here was also similar to the conclusions achieved by Nabiollahi et al. [30], who successfully used RF to map SOC stocks at the two depth intervals (0–30 and 30–60 cm) using RS- and terrain-based covariates, and found it performed fairly good to predict SOC at two soil depths ($R^2 = 0.70$ and 0.67 , respectively). However, Were et al. [20] input a wide range of the environmental covariates (e.g., soil properties, climate variables, land cover data, relief factors, and spectral indices) into RF and ANN to map SOC stocks, and showed that ANN had lower RMSE and ME values, as well as higher R^2 values in predicting SOC in comparison to RF models. The studies mentioned above have shown that the output of RF models varies significantly from study to study. Although, it is difficult to explain the reasons for these differences, but the differences could be because of the different extents of the study areas, topography, sampling densities, or quantity and quality of the environmental covariates used.

Table 5. Performances of the ML models for SOC content at six standard depths in the arid site.

Models	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ	R^2	RMSE	RPIQ	
		0–5 cm			5–15 cm			15–30 cm		
Cubist	0.76	0.25	0.84	0.63	0.24	0.75	0.63	0.20	0.67	
XGBoost	0.79	0.20	1.12	0.71	0.19	1.02	0.69	0.17	0.85	
RF	0.80	0.19	1.18	0.80	0.19	1.02	0.72	0.17	0.85	
ANN	0.75	0.19	1.05	0.67	0.19	0.89	0.65	0.16	0.78	
AvNNNet	0.78	0.20	1.06	0.69	0.18	1.01	0.66	0.17	0.79	
DNN	0.83	0.17	1.25	0.80	0.18	1.07	0.75	0.16	0.90	
Stack1	0.83	0.17	1.25	0.78	0.18	1.07	0.74	0.15	0.92	
Stack2	0.83	0.17	1.25	0.81	0.17	1.09	0.75	0.14	0.94	
Stack3	0.86	0.14	1.30	0.82	0.13	1.18	0.77	0.11	1.07	
Stack4	0.90	0.14	1.37	0.85	0.13	1.20	0.78	0.10	1.11	
		30–60 cm			60–100 cm			100–200 cm		
Cubist	0.49	0.14	0.92	0.29	0.13	0.90	0.17	0.16	0.78	
XGBoost	0.56	0.14	1.00	0.33	0.13	0.99	0.26	0.16	0.84	
RF	0.57	0.14	1.00	0.35	0.13	0.99	0.29	0.16	0.84	
ANN	0.50	0.13	0.91	0.29	0.11	0.97	0.22	0.15	0.77	
AvNNNet	0.53	0.14	0.92	0.31	0.12	0.98	0.24	0.15	0.83	
DNN	0.64	0.13	1.08	0.40	0.13	0.99	0.39	0.14	0.90	
Stack1	0.63	0.11	1.13	0.41	0.12	0.99	0.40	0.13	0.94	
Stack2	0.62	0.11	1.12	0.38	0.11	1.02	0.39	0.13	0.94	
Stack3	0.67	0.10	1.20	0.43	0.09	1.15	0.42	0.11	0.98	
Stack4	0.72	0.09	1.29	0.46	0.08	1.19	0.44	0.10	1.06	

R^2 : coefficient of determination; RMSE: root mean square error; RPIQ: Ratio of Performance to Interquartile distance; Stack: refers to Figure 2.

The performance of XGBoost at both sites and six depth intervals closely followed the performance of RF (Tables 5 and 6). In terms of R^2 , RMSE, and RPIQ, it outperformed Cubist. In agreement with our findings, Tziachris et al. [104] reported the reasonable accuracy of XGBoost in comparison with RF models to predict SOC in Greece. The higher accuracy of XGBoost can be explained because its stochastic gradient, which improves the procedure, can reduce overfitting, and can improve the prediction accuracy [67]. Furthermore, XGBoost ensemble has been shown to be capable of handling noise data due to the use of a number of decision-based tree classifiers. There are several other examples of DSM experts who applied XGBoost and RF models successfully to predict soil nutrient

in Sub-Saharan Africa [105], soil properties in the United States [106], soil pH in China [107], soil properties at the global scale [108], and the depth to bedrock at the global scale [109].

Table 6. Performances of the ML models for SOC content at six standard depths in the sub-humid site.

Models	R ²	RMSE	RPIQ	R ²	RMSE	RPIQ	R ²	RMSE	RPIQ	
		0–5 cm			5–15 cm			15–30 cm		
Cubist	0.78	1.35	2.00	0.76	1.26	1.90	0.66	1.17	1.62	
XGBoost	0.78	1.28	2.08	0.76	1.23	1.92	0.66	1.10	1.69	
RF	0.78	1.25	2.11	0.76	1.18	1.98	0.66	1.06	1.73	
ANN	0.78	1.31	2.04	0.76	1.25	1.89	0.65	1.13	1.65	
AvNNet	0.79	1.30	2.08	0.77	1.24	1.93	0.67	1.12	1.69	
DNN	0.81	1.26	2.12	0.79	1.17	2.02	0.69	1.05	1.78	
Stack1	0.83	1.21	2.16	0.82	1.17	2.05	0.73	1.06	1.78	
Stack2	0.83	1.20	2.19	0.82	1.16	2.04	0.74	1.03	1.79	
Stack3	0.84	1.16	2.25	0.85	1.13	2.06	0.74	1.01	1.81	
Stack4	0.87	1.15	2.29	0.86	1.12	2.10	0.78	1.01	1.83	
		30–60 cm			60–100 cm			100–200 cm		
Cubist	0.52	0.99	1.46	0.32	1.19	1.07	0.23	1.22	1.11	
XGBoost	0.61	0.95	1.49	0.36	1.12	1.12	0.27	1.15	1.16	
RF	0.61	0.92	1.51	0.38	1.08	1.14	0.26	1.14	1.15	
ANN	0.57	0.97	1.46	0.33	1.16	1.08	0.24	1.18	1.13	
AvNNet	0.62	0.96	1.50	0.36	1.15	1.11	0.28	1.16	1.17	
DNN	0.66	0.93	1.52	0.54	1.09	1.15	0.44	1.08	1.24	
Stack1	0.72	0.91	1.57	0.55	1.06	1.20	0.47	1.04	1.29	
Stack2	0.70	0.89	1.58	0.54	1.06	1.18	0.49	1.02	1.29	
Stack3	0.71	0.86	1.59	0.60	1.00	1.22	0.51	0.97	1.34	
Stack4	0.74	0.85	1.61	0.60	0.97	1.27	0.54	0.97	1.36	

R²: coefficient of determination; RMSE: root mean square error; RPIQ: Ratio of Performance to Interquartile distance; Stack: refers to Figure 2.

Although Cubist resulted in relatively good predictions of SOC content at the two study sites, especially at the surface layers (Tables 5 and 6), it was outperformed by RF and XGBoost. In line with our results, Zeraatpisheh et al. [21] revealed that, in terms of R² and RMSE, Cubist was outperformed by RF. Despite, the usefulness of Cubist in explaining the relationships between soil properties and covariates and in modeling SOC content, which has been reported in several studies [93], our findings showed that the model was not very competitive with the other ML models. It is not clear why Cubist failed to produce higher accuracies in comparison to the other ML models in the current research. The different findings could be related to differences in the different processes that cause the evolution and accumulation of SOC in these soils. Furthermore, this suggests that no single ML algorithm could best serve for every landscape and that multiple models should be optimized to find the most reliable prediction model. Nevertheless, we noted that the differences in the ML models' performance at both sites and at all depth intervals were rather small.

For the two areas, the performance of classical ANN at all depth intervals closely followed the performance of AvNNet (Tables 5 and 6). The higher performance of AvNNet, compared to ANN, was also reported by Baker and Ellison [84], who evaluated and compared the performance of AvNNet and ANN in order to predict water retention data. They indicated that combining ANNs improves the ability to generalize individual component ANNs. Similarly, Meyer et al. [83] reported the higher performance of AvNNet by comparing it with three ML models—RF, ANN, and SVM—for rainfall area detection and rainfall rate assignment over Germany. Although the performance of prediction slightly improves by averaging ANNs in comparison to the classical ANN, Meyer et al. [83] concluded that predictions might have been advantageous in cases where only limited data are available for training. Our result, however, is different from that of Taghizadeh-Mehrjardi et al. [7], who found superior performance for ANNs compared to RF for the three-dimensional mapping of SOC content in the western parts of Iran. It should be noted that the prediction accuracy can be affected by several parameters, such as number of field observation, type of models, the variability of soil properties, and the ability of environmental covariates to describe SOC variations.

In the arid site, the DNN was able to account for 39% to 83% of the total variation of SOC content from the lower depth interval (100 to 200 cm) to the surface layer (0–5 cm), respectively (Table 5). For the sub-humid site (Table 6), the DNN showed the best performance. It was able to account for 44% to 81% of total SOC content variation at the depths of 100–200 cm and 0–5 cm, respectively. Overall, DNN outperformed the other individual methods and produced the most accurate results, though all five ML models in the study areas proved appropriate for SOC mapping. It can be concluded that DNN provided the best outcomes in this study compared with the other individual methods and demonstrated the ML's robustness in complex data modeling and prediction. Our results are in line with other studies in the ML literature that reported the capability of the DNN model to reveal and learn the non-linear and complex patterns underlain datasets [32,110]. In soil science literature, however, the superior performance of DNN in predicting soil properties is only reported in a few studies [32,111]. For instance, Behrens et al. [31] found the most accurate results for DSM analysis using deep learning, indicating an improvement of 4–7% compared to RF. Additionally to this example, Padarian et al. [33] and Wadoux et al. [34] successfully applied a convolutional neural network (CNN) model (a well-known DNN model) to predict different soil properties (e.g., SOC) from large spectroscopic databases. Similar to our results, they also achieved a better performance by implementing a CNN compared to other individual ML models. This is mainly because of the fact that the CNN models use the contextual covariates information as input [34]. Nevertheless, our proposed DNN used the point covariates as input in this research. We can, therefore, conclude that the higher DNN output compared to the other individual ML methods can be related to the other factors, including the power of features' or attributes' extraction from raw data. DNN models create new features (sometimes also called representations of the raw data) automatically using neural networks with many hidden layers and powerful computational resources to allow them to model highly complex functions, e.g., SOC–landscape relationships.

3.4. Performances of the Stacking Ensemble Models

The modeling results of the two stacking approaches, namely, the standard mode (Stack1 and Stack2) and the rescan mode (Stack3 and Stack4), for the prediction of SOC content at the six depth intervals in the arid and sub-humid areas are presented in Tables 5 and 6. The modeling accuracy is as follows: Stack4 > Stack3 > Stack2 > Stack1. For instance, the RMSEs for the SOC content prediction in the depth interval of 0 to 5 cm of both sites by Stack4 models were 17% and 5% lower than the ones obtained by Stack1 (Tables 5 and 6). Similarly, Stack4 reduced the RMSEs (23% and 6%) for SOC content prediction in 100 to 200 cm of soil profiles of the arid site and the sub-humid site, compared to the Stack1 models. Generally, stacking ensemble models in rescan mode had higher accuracy than the ones in the standard mode [68]. This might be related to the fact that the meta-learning models (LASSO and SVR) used in level 1 can recapture and extract some missing information from the original covariate space. This proves the potential of using rescanning the original covariate space to improve the performance of standard stacking methods. Note that there was no attempt to test the other ML algorithms at level 0 and level 1. By changing the number and type of individual and meta-learning models, one might possibly further improve the performance of the proposed stacking method. In addition, prediction accuracy can be further improved if the model residuals' spatial correlation structure is analyzed and then added to the determinist spatial trend. However, such an analysis was beyond the scope of our study.

In order to further understand which ML models performed the best, we illustrated the performances of the stacking and individual models in Figure 6. The graph shows that the stacking ensemble modeling in both modes (standard mode and rescan mode) indicated the higher performance in comparison to the individual models. Here, we compare the performances of the best individual model (DNN) and the best stacking model (Stack4), in terms of R^2 and RMSE. What can be clearly seen in Figure 6 is that the Stack4 ensemble models increased R^2 values and decreased the RMSE values in comparison to the DNN models. For instance, the RMSEs for SOC content prediction in 0 to 5 cm of

soil profiles of the arid site and sub-humid site by Stack4 models were 17% and 9% respectively, less than that obtained by the DNN models. Similarly, Stack4 reduced the RMSE values (28% and 10%) for SOC content prediction in 100 to 200 cm of soil profiles of the arid site and sub-humid site, compared to the DNN models.

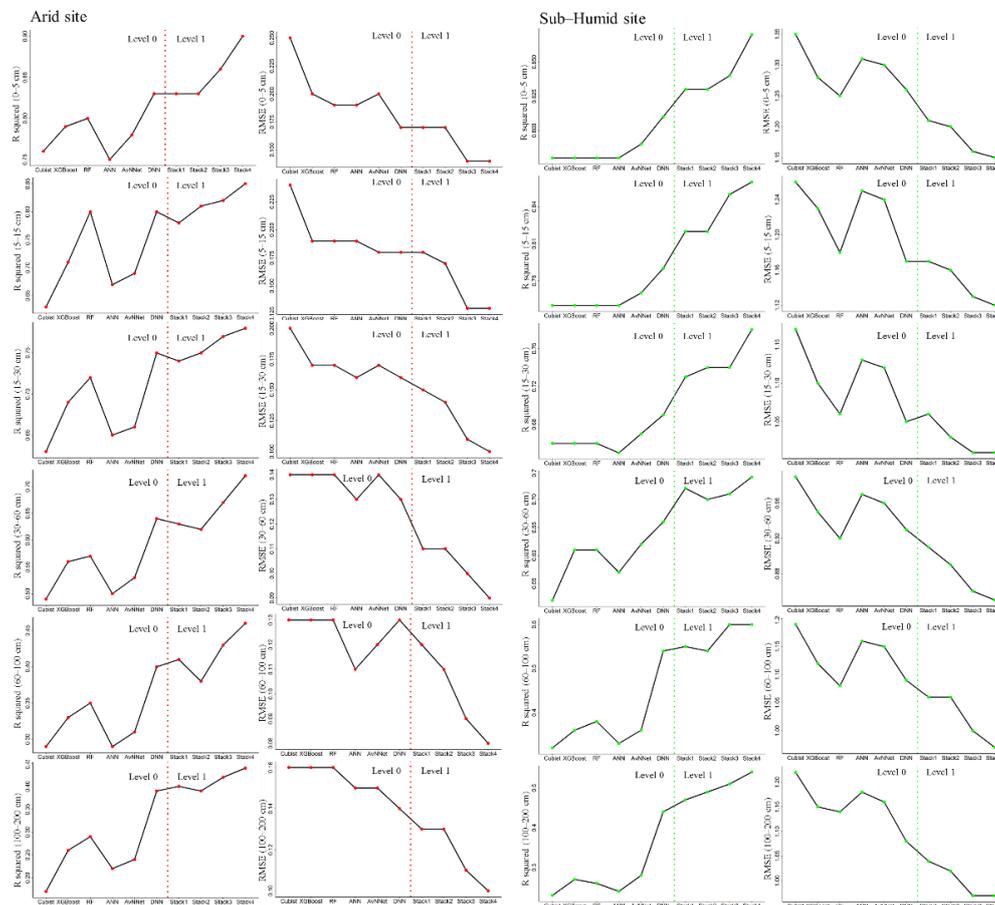


Figure 6. Coefficient of determination (R^2) and root mean squared error (RMSE) values of the individual models (Level 0) and stacking models (Level 1) for SOC content at six standard depths in two regions.

Generally, Stack4 ensemble models exhibited the best competence for capturing the spatial variation of SOC content and reducing prediction uncertainty as well. This indicated that the stacking ensemble models in level 1 were successful to keep the advantages and to discard the inaccurate aspect of the individual ML models in level 0. This is justified by the fact that the information lost by the models in level 0 is successfully captured by the level 1 models. In fact, the stacking methods used multiple learning algorithms' strengths to obtain better predictive performance and make the predictive model more robust than it is from the individual models. We emphasize that the stacking strategy is more accurate than any of its individual models if the individual models are accurate and diverse. The success of the stacking method would generally be linked to two facts: (1) the training data does not always provide enough information to select a single accurate model, and (2) the learning processes of the individual model may be imperfect [40,68]. Those are the reasons why stacking never did worse than selecting the individual models in our case study. Several studies have revealed that ensemble models exhibited the best performances for predicting soil properties in the DSM community [38,68]. For instance, Tajik et al. [38] found that stack modeling showed better performance to predict SOC content in comparison to the individual models, including RF and SVM. Similarly, Zhou et al. [39] and Chen et al. [40] recently evaluated the efficacy of the ensemble models to predict the spatial variation of soil properties. In order to predict soil total nitrogen, Zhou et al. [39] combined RF and XGBoost models

using the weighted averaging method. Their results confirmed a reasonable outcome was obtained by the ensemble approach with the lowest RMSE ($1.15 \text{ g}\cdot\text{kg}^{-1}$) and the highest R^2 (0.41) compared with the two individual models. Somarathna et al. [112], however, concluded that combining the ML algorithms could not provide a significant improvement ($\sim 2\%$) in SOC predictions. They reported that the ensemble model can only prove beneficial when combining different ML algorithms, covariates or datasets. This implies stacking strategies tend to produce reasonable performance depending on the diversification of the individual models. It can be mentioned in the current research that a diverse and powerful set of ML algorithms implemented in level 0, resulting in the stacking methods, suggested higher prediction performance compared to the individual models. Added to this, the rescanning of the original covariate space in level 1 increased the variety and strength of stacking methods, proving to be an effective tool for predicting SOC contents, which should be used as a tool in digital soil mapping.

Although the effectiveness of stacking has been demonstrated in several studies (e.g., References [38–40]), there are some potential drawbacks that should be considered. Since the most improvement in stacking is obtained when the predictors at level 0 are less correlated, selecting a combination of dissimilar models is not always an easy practice. This technique is also typically expensive in computational terms. Therefore, they add learning time and memory constraints to the problem. Last but not the least, because the stacking is a “black box” algorithm, the exact contributions of predictors to the final output cannot be explicitly disclosed. This means that the stacking (ensemble) models suffer from a lack of interpretability.

Several tools have been developed to gain insight into the fitted function and increase the interpretability of ML models, e.g., partial dependence plot (PDP) and permutation variable importance [113]. Partial plots are an intuitive and easy-to-understand visualizations technique to show the marginal effects of each predictor on the response [113,114]. However, there are limitations in using PDPs, for instance, the PDPs are meaningful when the input covariates are not highly correlated [113] and the response curve does not consider the interaction between the covariates and is more meaningful for additive models [114]. The input **feature** of stacked models (i.e., the rescanned original covariate space and the prediction of our six models) are both highly correlated (high correlation between the prediction of different models) and have a high level of interaction (as the prediction of each model directly depends on the rescanned original covariate space). Thus, using PDPs might not be very reasonable in assessing the fitted function of stacked models. This is one of the disadvantages of using model stacking as this technique makes the interpretation of models exceedingly difficult. Adding the rescan covariate space also causes further difficulties in model interpretation. In general, the interpretability of ML models in real-world problems, such as digital soil mapping, is still a challenge that needs to be focused on in the future.

3.5. Performances of ML Models in Two Different Climatic Regions

Vertical distribution of R^2 , n RMSE, and RPIQ values to a depth of 200 cm are depicted in Figure 7. Generally, the two sites showed a decreasing trend in R^2 and RPIQ values with increasing depth. Otherwise, the percentage of variation in SOC content, which is described by the models, decreased with increasing depth. A reverse trend for n RMSE was also revealed (Figure 7). Results indicate that the models' performance decreased by each depth increment down the soil profile, and confirmed that the models had much better prediction efficiency for surface layers than subsurface layers. Increasing uncertainty of SOC content with depth has been reported in numerous studies [24]. For instance, Laub et al. [115] found that the efficiency of the ML models used for SOC prediction in China decreased from about 0.8 in the topsoil to 0.2 at 0.8 to 1 m subsoil depth. A similar pattern of uncertainty variation with depth was reported in several other studies [22]. This decreasing trend in performances could be explained by the fact that most of terrain- and RS-based covariates that are used as predictors of SOC content (listed in Table 2) explain soil surface features and processes [9] and, for example, vertical hydrological processes or bioturbation are less explained; therefore, the covariates used cannot capture the subsurface SOC variation.

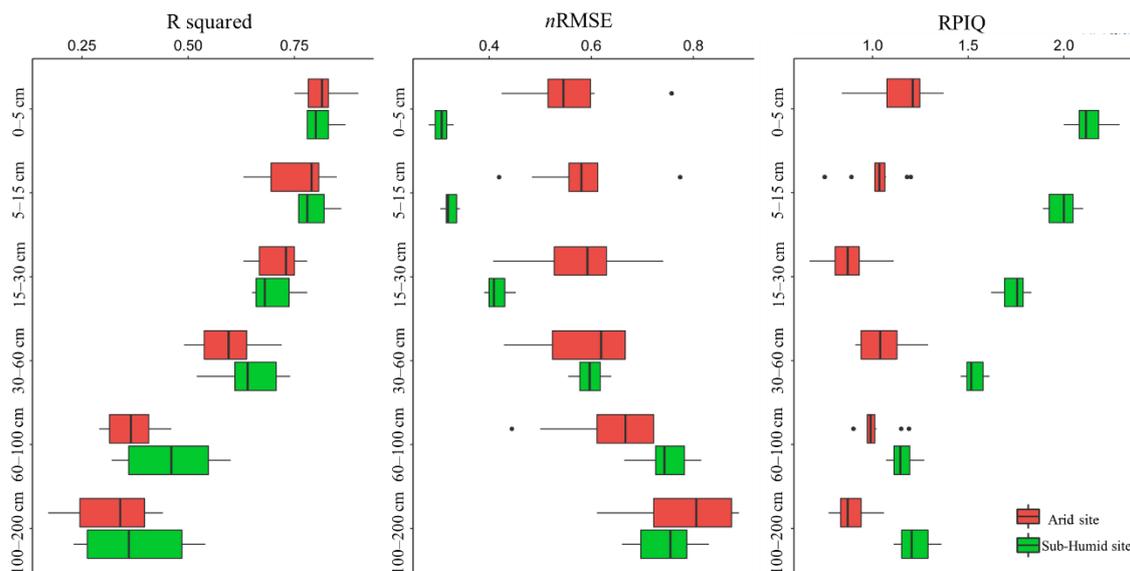


Figure 7. Comparison of prediction power of ML models for SOC content at six standard depths in two regions.

In line with all performance measures in Figure 7 (R^2 , $nRMSE$, and RPIQ), all tested ML models performed better for the six soil depths at the sub-humid site than for those of the arid site. Here, we compare the performances of the best (Stack4), the worst (Cubist), and AvNNNet as a model with intermediate performance, in terms of R^2 , $nRMSE$, and RPIQ.

Stack4 resulted in R^2 values on average of 0.73 and 0.69 for the sub-humid site and arid site, respectively. R^2 values for Cubist were 0.55 and 0.50 and for AvNNNet 0.58 and 0.54 for the same areas as for Stack4. Our results, furthermore, indicated that Stack4, Cubist, and AvNNNet resulted in R^2 values for the sub-humid site that were ~9%, ~8%, and ~5% higher than the values of the arid site. The difference in the performance of ML models is much more evident when we consider $nRMSE$ values, in which $nRMSE$ values obtained by Stack4, Cubist, and AvNNNet for the arid site were ~25%, ~17%, and ~2% more than those values for the sub-humid site (Figure 7). These results indicated that the ML models performed better at six depth intervals in the sub-humid site in comparison to those obtained in the arid site. These results could be partly attributed to large differences between areas in terms of soil forming factors, which results in complex relationships between SOC content and covariates.

The ML models resulted in a decreasing and increasing trend in R^2 , $nRMSE$, and RPIQ respectively, with depth in the two areas (Figure 7). As can be seen, at the top layers (0–60 cm), the arid site tended to have the highest values of $nRMSE$, but with increasing depth, the accuracy of models in terms of $nRMSE$ tended to be almost the same at both sites. This further shows that ML models, such as Stack4 based on the covariates used in this study, cannot capture SOC content variability at the bottom of soil profiles [7,54]. This is consistent with the results of other researchers [7,17,22,54] who all reported that the accuracy of DSM decreased with increasing depth. This indicates that the biggest uncertainty is driven by covariate space, not by the selection of an ML algorithm, which may be due to the reduction of the explanatory power of the environmental covariates along with the increase of the depth of the soil layer. It should be added that the typical covariates used in DSM studies are RS- and terrain-based covariates [10], which can only characterize the soil properties at the earth's surface. The results of this study show that there is still work to be done to improve the model accuracy for predictions of SOC in depth. Moreover, the model accuracy could potentially be improved through the use of additional environmental covariates. We assume that an implementation of important terrain-based scale-relevant covariates representing climate-driven proxy variables (e.g., shadowing effects of high mountain ranges) or the generation of covariates representing the vertical distribution of water, will influence the prediction accuracies in deeper horizons (e.g., using the geophysical techniques).

3.6. Spatial Distribution of SOC

The spatial distribution of mean, upper, and lower limits of SOC contents at the arid site at six interval depths is depicted in Figure 8. A decreasing trend in SOC content down the soil profile was observed. Central parts of the area tended to have the largest amounts of SOC content, which correspond to the cultivated areas mainly under pistachio orchards and wheat. Moreover, the topography of this area is mainly flat and located downslope, which results in more accumulation of fine-textured materials and water. In the arid site, because of rainfall scarcity, irrigation is necessary to provide soil moisture for crop production. Thus, irrigated farming and topographic attributes in the central parts promote more vegetation and consequently, more organic matter is accumulated in the soil. Lower SOC content in the other parts of the area can be attributed to the higher slope degree, which makes these areas prone to erosion and to higher water discharge. Further, water scarcity in these areas is not compensated by irrigation. In line with our results, Wiesmeier et al. [52] suggested that at the small scales with similar climatic conditions, vegetation, land use, and land management have a significant influence on the level of SOC stocks. At the regional scale, climatic effects may be counterbalanced by agricultural practices (e.g., fertilization and irrigation) [49].

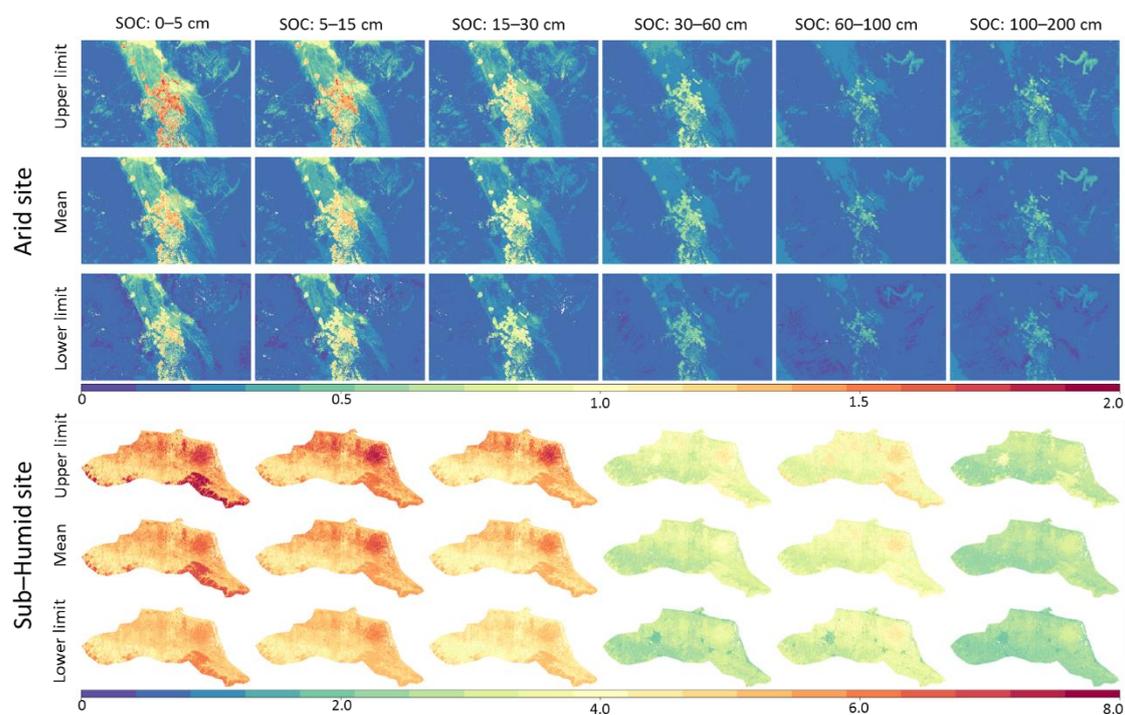


Figure 8. The spatial distribution of mean, upper, and lower limits of SOC contents at six standard depths in two regions. The upper and lower limits were calculated using $\text{Mean} \pm (1.5 \times \text{standard deviation (SD)})$ of the prediction values of SOC contents using the spatial block cross-validation.

The spatial distribution of mean, upper, and lower limits of SOC contents at the sub-humid site at six interval depths is depicted in Figure 8. Again, a decreasing trend in SOC content with depth is shown in the sub-humid site. The map of the spatial distribution of SOC content in the upper layer revealed more SOC accumulation in the northern parts than the other sections. The low slope degrees of the northern parts make these areas favorable for more water accumulation and, thus, result in poorly drained soils. SOC content is more accumulated and less decomposed in poorly drained soils. Mishra et al. [116] reported that high SOC stocks were found in areas characterized by low slope gradient and poorly drained soils. Wiesmeier et al. [52] indicated that areas with low slope degree and concave surface favor water accumulation. Soil moisture, which is largely controlled by terrain attributes, affects the spatial distribution of SOC content [117].

4. Conclusions

In this study, we introduced stacking ML models in two modes (standard mode and rescan mode) in order to improve the spatial prediction of SOC content at two contrasting climatic regions (arid and sub-humid) of Iran. The main conclusions are:

1. Though the differences in the ML models' performance at both sites and at all depth intervals were rather small, DNN was identified as the most suitable individual model.
2. The stacking ensemble modeling in both modes (standard mode and rescan mode) indicated the higher performance in comparison to the individual models.
3. Although both terrain- and RS-based covariates were important to explain SOC contents at both sites, their explanatory power was different at both sites and at the soil depth intervals.
4. The stacking models are able to explain the effect of contrasting climate on SOC content distribution. Higher content of SOC in the sub-humid site and lower content of SOC in the arid site were found, however local variation is controlled by moisture, terrain, and land use.

Author Contributions: Conceptualization—R.T.-M., T.B., T.S., and K.S.; methodology—R.T.-M., T.R., R.V., T.B., and K.S.; software—R.T.-M.; analysis—R.T.-M. and K.S.; investigation—R.T.-M., A.A.-C., T.B., T.S., and K.S.; data curation—R.T.-M., F.S., and N.D.; writing—original draft preparation—R.T.-M., A.A.-C., T.R., M.Z., R.V., F.S., and K.S.; visualization—R.T.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: R.T.-M. has been supported by the Alexander von Humboldt Foundation (grant number: Ref3.4-1164573-IRN-GFHERMES-P). K.S., T.R., and T.S. thank the German Research Foundation (DFG) for supporting this research through the Collaborative Research Center (SFB 1070) 'ResourceCultures' (subprojects Z, S and B02). K.S., T.B., and T.S. are also supported by the DFG Cluster of Excellence "Machine Learning—New Perspectives for Science", EXC 2064/1, project number 390727645. R.V. is supported by an Australian Government Research Training Program Scholarship and a Rowden White Scholarship. Sandra Teuber, Department of Geosciences, University of Tübingen, Tübingen, Germany, thoroughly revised technical English of the paper. We acknowledge support by Open Access Publishing Fund of University of Tübingen. Finally, we thank the anonymous reviewers and editors for their careful reading of our manuscript and their many insightful comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

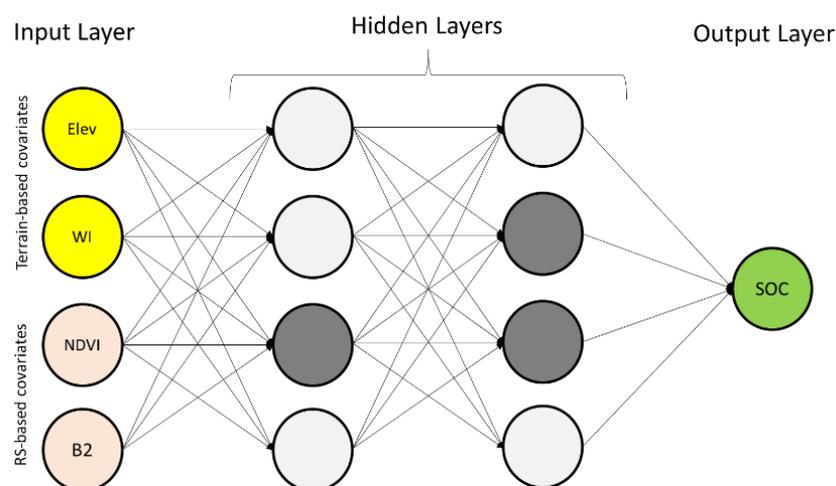


Figure A1. Illustration of the DNN model with two hidden layers. Black-colored circles are neurons deactivated using the dropout method. (Elev: elevation; WI: wetness index; NDVI: normalized difference vegetation index; B2: the second band of remote sensed data; SOC: soil organic carbon content).

Table A1. The optimal set of hyper-parameters used for ML algorithms for prediction of SOC in the arid site.

ML Models	Hyper-Parameters	Arid Site					
		SOC 0–5 cm	SOC 5–15 cm	SOC 15–30 cm	SOC 30–65 cm	SOC 60–100 cm	SOC 100–200 cm
Cubist	committees	3	3	7	5	4	3
	neighbors	4	3	4	4	7	2
XGboost	booster	gbtree	gbtree	gbtree	gbtree	gbtree	gbtree
	max_depth	6	4	7	6	5	6
	min_child_weight	2	1	2	1	3	1
	colsample_bytree	0.5	0.5	0.5	0.5	0.5	0.5
	subsample	0.75	0.75	0.5	0.75	0.25	0.5
	eta	0.3	0.3	0.2	0.2	0.3	0.3
RF	Mtry	9	11	12	18	16	22
	Ntree	800	500	1100	1200	1800	2400
ANN	decay	0.01	0.01	0.03	0.03	0.03	0.01
	size	8	5	6	5	8	8
AvNNNet	Repeats	14	10	9	18	24	7
DNN	Hidden	4	4	6	5	6	8
	Size	15	20	30	40	30	50
	Network weight initialization	uniform	uniform	uniform	uniform	uniform	uniform
	learning rate	0.02	0.05	0.01	0.03	0.01	0.02
	dropout regularization	0.7	0.6	0.3	0.4	0.4	0.8

Table A2. The optimal set of hyper-parameters used for ML algorithms for prediction of SOC in the sub-humid site.

ML Models	Hyper-Parameters	Sub-Humid Site					
		SOC 0–5 cm	SOC 5–15 cm	SOC 15–30 cm	SOC 30–65 cm	SOC 60–100 cm	SOC 100–200 cm
Cubist	Committees	4	5	3	8	7	5
	neighbors	5	3	2	2	7	8
XGboost	booster	gbtree	gbtree	gbtree	gbtree	gbtree	gbtree
	max_depth	6	5	6	5	6	4
	min_child_weight	2	1	1	4	3	2
	colsample_bytree	0.5	0.5	0.5	0.5	0.5	0.5
	subsample	0.5	0.5	0.5	0.75	0.5	0.5
	eta	0.3	0.3	0.2	0.2	0.3	0.4
RF	Mtry	14	11	17	16	21	24
	Ntree	1400	900	1600	2100	2600	1900
ANN	decay	0.01	0.01	0.03	0.03	0.03	0.01
	size	8	5	6	5	8	8
AvNNNet	Repeats	14	10	9	18	24	7
DNN	hidden	4	4	6	5	6	8
	size	50	20	40	40	50	60
	Network weight initialization	uniform	uniform	uniform	uniform	uniform	uniform
	learning rate	0.02	0.05	0.01	0.03	0.01	0.02
	dropout regularization	0.7	0.6	0.3	0.4	0.4	0.8

References

- Scholten, T.; Goebes, P.; Kühn, P.; Seitz, S.; Assmann, T.; Bauhus, J.; Bruelheide, H.; Buscot, F.; Erfmeier, A.; Fischer, M. On the combined effect of soil fertility and topography on tree growth in subtropical forest ecosystems—A study from SE China. *J. Plant Ecol.* **2017**, *10*, 111–127. [\[CrossRef\]](#)
- Minasny, B.; McBratney, A. Limited effect of organic matter on soil available water capacity. *Eur. J. Soil Sci.* **2018**, *69*, 39–47. [\[CrossRef\]](#)
- Don, A.; Schumacher, J.; Scherer-Lorenzen, M.; Scholten, T.; Schulze, E.-D. Spatial and vertical variation of soil carbon at two grassland sites—implications for measuring soil carbon stocks. *Geoderma* **2007**, *141*, 272–282. [\[CrossRef\]](#)
- Adhikari, K.; Owens, P.R.; Libohova, Z.; Miller, D.M.; Wills, S.A.; Nemecek, J. Assessing soil organic carbon stock of Wisconsin, USA and its fate under future land use and climate change. *Sci. Total Environ.* **2019**, *667*, 833–845. [\[CrossRef\]](#)

5. Minasny, B.; McBratney, A.; Malone, B.; Wheeler, I. Digital soil mapping of carbon. *Adv. Agron.* **2013**, *118*, 1–47.
6. Ajami, M.; Heidari, A.; Khormali, F.; Gorji, M.; Ayoubi, S. Environmental factors controlling soil organic carbon storage in loess soils of a subhumid region, northern Iran. *Geoderma* **2016**, *281*, 1–10. [[CrossRef](#)]
7. Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Kerry, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* **2016**, *266*, 98–110. [[CrossRef](#)]
8. Bernhard, N.; Moskwa, L.-M.; Schmidt, K.; Oeser, R.A.; Aburto, F.; Bader, M.Y.; Baumann, K.; von Blanckenburg, F.; Boy, J.; van den Brink, L. Pedogenic and microbial interrelations to regional climate and local topography: New insights from a climate gradient (arid to humid) along the Coastal Cordillera of Chile. *Catena* **2018**, *170*, 335–355. [[CrossRef](#)]
9. Zeraatpisheh, M.; Jafari, A.; Bodaghabadi, M.B.; Ayoubi, S.; Taghizadeh-Mehrjardi, R.; Toomanian, N.; Kerry, R.; Xu, M. Conventional and digital soil mapping in Iran: Past, present, and future. *Catena* **2020**, *188*, 104424. [[CrossRef](#)]
10. Ma, Y.; Minasny, B.; Malone, B.P.; McBratney, A.B. Pedology and digital soil mapping (DSM). *Eur. J. Soil Sci.* **2019**, *70*, 216–235. [[CrossRef](#)]
11. McBratney, A.B.; Stockmann, U.; Angers, D.A.; Minasny, B.; Field, D.J. Challenges for soil organic carbon research. In *Soil Carbon*; Springer: Berlin, Germany, 2014; pp. 3–16.
12. Minasny, B.; McBratney, A.B.; Lark, R.M. Digital soil mapping technologies for countries with sparse data infrastructures. In *Digital Soil Mapping with Limited Data*; Springer: Berlin, Germany, 2008; pp. 15–30.
13. Behrens, T.; Scholten, T. Digital soil mapping in Germany—A review. *J. Plant Nutr. Soil Sci.* **2006**, *169*, 434–443. [[CrossRef](#)]
14. McBratney, A.B.; Santos, M.M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [[CrossRef](#)]
15. Grimm, R.; Behrens, T.; Märker, M.; Elsenbeer, H. Soil organic carbon concentrations and stocks on Barro Colorado Island—Digital soil mapping using Random Forests analysis. *Geoderma* **2008**, *146*, 102–113. [[CrossRef](#)]
16. Moreno, R.; Irigoyen, A.I.; Monterubbianesi, M.G.; Studdert, G.A. Application of artificial neural networks to estimate soil organic carbon in a high-organic-matter Mollisol. *Span. J. Soil Sci. SJSS* **2017**, *7*, 179–200.
17. Rentschler, T.; Gries, P.; Behrens, T.; Bruelheide, H.; Kühn, P.; Seitz, S.; Shi, X.; Trogisch, S.; Scholten, T.; Schmidt, K. Comparison of catchment scale 3D and 2.5 D modeling of soil organic carbon stocks in Jiangxi Province, PR China. *PLoS ONE* **2019**, *14*, e0220881. [[CrossRef](#)] [[PubMed](#)]
18. Stumpf, F.; Keller, A.; Schmidt, K.; Mayr, A.; Gubler, A.; Schaeppman, M. Spatio-temporal land use dynamics and soil organic carbon in Swiss agroecosystems. *Agric. Ecosyst. Environ.* **2018**, *258*, 129–142. [[CrossRef](#)]
19. Wang, B.; Waters, C.; Orgill, S.; Cowie, A.; Clark, A.; Li Liu, D.; Simpson, M.; McGowen, I.; Sides, T. Estimating soil organic carbon stocks using different modeling techniques in the semi-arid rangelands of eastern Australia. *Ecol. Indic.* **2018**, *88*, 425–438. [[CrossRef](#)]
20. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. 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. *Ecol. Indic.* **2015**, *52*, 394–403. [[CrossRef](#)]
21. Zeraatpisheh, M.; Ayoubi, S.; Jafari, A.; Tajik, S.; Finke, P. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* **2019**, *338*, 445–452. [[CrossRef](#)]
22. Amirian-Chakan, A.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Kumar, S.; Khordehbin, S.; Khanghah, S.Y. Spatial 3D distribution of soil organic carbon under different land use types. *Environ. Monit. Assess.* **2017**, *189*, 131. [[CrossRef](#)]
23. Owusu, S.; Yigini, Y.; Olmedo, G.F.; Omuto, C.T. Spatial prediction of soil organic carbon stocks in Ghana using legacy data. *Geoderma* **2020**, *360*, 114008. [[CrossRef](#)]
24. Adhikari, K.; Hartemink, A.E.; Minasny, B.; Kheir, R.B.; Greve, M.B.; Greve, M.H. Digital mapping of soil organic carbon contents and stocks in Denmark. *PLoS ONE* **2014**, *9*, e105519. [[CrossRef](#)] [[PubMed](#)]
25. Hinge, G.; Surampalli, R.Y.; Goyal, M.K. Prediction of soil organic carbon stock using digital mapping approach in humid India. *Environ. Earth Sci.* **2018**, *77*, 172. [[CrossRef](#)]
26. Dharumarajan, S.; Kalaiselvi, B.; Suputhra, A.; Lalitha, M.; Hegde, R.; Singh, S.; Lagacherie, P. Digital soil mapping of key GlobalSoilMap properties in Northern Karnataka Plateau. *Geoderma Reg.* **2020**, *20*, e00250. [[CrossRef](#)]

27. Zhao, Z.; Yang, Q.; Sun, D.; Ding, X.; Meng, F.-R. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. *Comput. Electron. Agric.* **2020**, *169*, 105172. [[CrossRef](#)]
28. Guo, L.; Fu, P.; Shi, T.; Chen, Y.; Zhang, H.; Meng, R.; Wang, S. Mapping field-scale soil organic carbon with unmanned aircraft system-acquired time series multispectral images. *Soil Tillage Res.* **2020**, *196*, 104477. [[CrossRef](#)]
29. Yang, L.; He, X.; Shen, F.; Zhou, C.; Zhu, A.-X.; Gao, B.; Chen, Z.; Li, M. Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data. *Soil Tillage Res.* **2020**, *196*, 104465. [[CrossRef](#)]
30. Nabiollahi, K.; Eskandari, S.; Taghizadeh-Mehrjardi, R.; Kerry, R.; Triantafalis, J. Assessing soil organic carbon stocks under land-use change scenarios using random forest models. *Carbon Manag.* **2019**, *10*, 63–77. [[CrossRef](#)]
31. Behrens, T.; Schmidt, K.; MacMillan, R.A.; Rossel, R.A.V. Multi-scale digital soil mapping with deep learning. *Sci. Rep.* **2018**, *8*, 15244. [[CrossRef](#)]
32. Ng, W.; Minasny, B.; Montazerolghaem, M.; Padarian, J.; Ferguson, R.; Bailey, S.; McBratney, A.B. Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra. *Geoderma* **2019**, *352*, 251–267. [[CrossRef](#)]
33. Padarian, J.; Minasny, B. Using deep learning for digital soil mapping. *Soil* **2019**, *5*, 79–89. [[CrossRef](#)]
34. Wadoux, A.M.-C.; Padarian, J.; Minasny, B. Multi-source data integration for soil mapping using deep learning. *Soil* **2019**, *5*, 107–119. [[CrossRef](#)]
35. Breiman, L. Stacked regressions. *Mach. Learn.* **1996**, *24*, 49–64. [[CrossRef](#)]
36. Caubet, M.; Dobarco, M.R.; Arrouays, D.; Minasny, B.; Saby, N.P. Merging country, continental and global predictions of soil texture: Lessons from ensemble modeling in France. *Geoderma* **2019**, *337*, 99–110. [[CrossRef](#)]
37. Malone, B.P.; Minasny, B.; Odgers, N.P.; McBratney, A.B. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* **2014**, *232*, 34–44. [[CrossRef](#)]
38. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [[CrossRef](#)]
39. Ribeiro, M.H.D.M.; dos Santos Coelho, L. Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series. *Appl. Soft Comput.* **2020**, *86*, 105837. [[CrossRef](#)]
40. Tajik, S.; Ayoubi, S.; Zeraatpisheh, M. Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran. *Geoderma Reg.* **2020**, *20*, e00256. [[CrossRef](#)]
41. Zhou, Y.; Xue, J.; Chen, S.; Zhou, Y.; Liang, Z.; Wang, N.; Shi, Z. Fine-Resolution Mapping of Soil Total Nitrogen across China Based on Weighted Model Averaging. *Remote Sens.* **2020**, *12*, 85. [[CrossRef](#)]
42. Chen, S.; Mulder, V.L.; Heuvelink, G.B.; Poggio, L.; Caubet, M.; Dobarco, M.R.; Walter, C.; Arrouays, D. Model averaging for mapping topsoil organic carbon in France. *Geoderma* **2020**, *366*, 114237. [[CrossRef](#)]
43. Smith, D. *Soil Survey Staff: Keys to Soil Taxonomy*; Natural Resources Conservation Service: Washington, DC, USA, 2014.
44. Taghizadeh-Mehrjardi, R.; Minasny, B.; Sarmadian, F.; Malone, B. Digital mapping of soil salinity in Ardakan region, central Iran. *Geoderma* **2014**, *213*, 15–28. [[CrossRef](#)]
45. Minasny, B.; McBratney, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* **2006**, *32*, 1378–1388. [[CrossRef](#)]
46. Stumpf, F.; Schmidt, K.; Behrens, T.; Schönbrodt-Stitt, S.; Buzzo, G.; Dumperth, C.; Wadoux, A.; Xiang, W.; Scholten, T. Incorporating limited field operability and legacy soil samples in a hypercube sampling design for digital soil mapping. *J. Plant Nutr. Soil Sci.* **2016**, *179*, 499–509. [[CrossRef](#)]
47. Gholizadeh, A.; Zizala, D.; Saberioon, M.; Boruvka, L. Soil organic carbon content monitoring and mapping using airborne and Sentinel-2 spectral imaging. In Proceedings of the Sixth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2018), Paphos, Cyprus, 26–29 March 2018; p. 1077319.
48. Gholizadeh, A.; Žizala, D.; Saberioon, M.; Boruvka, L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. *Remote Sens. Environ.* **2018**, *218*, 89–103. [[CrossRef](#)]
49. Goidts, E.; Wesemael, B.V.; Van Oost, K. Driving forces of soil organic carbon evolution at the landscape and regional scale using data from a stratified soil monitoring. *Glob. Chang. Biol.* **2009**, *15*, 2981–3000. [[CrossRef](#)]

50. Roozitalab, M.H.; Toomanian, N.; Dehkordi, V.R.G.; Khormali, F. Major soils, properties, and classification. In *The Soils of Iran*; Springer: Berlin, Germany, 2018; pp. 93–147.
51. Taghizadeh-Mehrjardi, R.; Neupane, R.; Sood, K.; Kumar, S. Artificial bee colony feature selection algorithm combined with machine learning algorithms to predict vertical and lateral distribution of soil organic matter in South Dakota, USA. *Carbon Manag.* **2017**, *8*, 277–291. [[CrossRef](#)]
52. Wiesmeier, M.; Urbanski, L.; Hobbey, E.; Lang, B.; von Lützw, M.; Marin-Spiotta, E.; van Wesemael, B.; Rabot, E.; Ließ, M.; Garcia-Franco, N. Soil organic carbon storage as a key function of soils—a review of drivers and indicators at various scales. *Geoderma* **2019**, *333*, 149–162. [[CrossRef](#)]
53. Nelson, D.W.; Sommers, L.E. Total carbon, organic carbon, and organic matter. *Methods Soil Anal. Part 3 Chem. Methods* **1996**, *5*, 961–1010.
54. Malone, B.P.; McBratney, A.; Minasny, B.; Laslett, G. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **2009**, *154*, 138–152. [[CrossRef](#)]
55. Arrouays, D.; Grundy, M.G.; Hartemink, A.E.; Hempel, J.W.; Heuvelink, G.B.; Hong, S.Y.; Lagacherie, P.; Lelyk, G.; McBratney, A.B.; McKenzie, N.J. GlobalSoilMap: Toward a fine-resolution global grid of soil properties. In *Advances in Agronomy*; Elsevier: Amsterdam, The Netherlands, 2014; Volume 125, pp. 93–134.
56. Xiong, X.; Grunwald, S.; Myers, D.B.; Ross, C.W.; Harris, W.G.; Comerford, N.B. Interaction effects of climate and land use/land cover change on soil organic carbon sequestration. *Sci. Total Environ.* **2014**, *493*, 974–982. [[CrossRef](#)]
57. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L. The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*, 1–33. [[CrossRef](#)]
58. Conrad, O.; Olaya, V. SAGA-GIS module library documentation (v2. 2.3). *Module Valley Depth*. Available online: http://www.saga-gis.org/saga_tool_doc/2.2.3/index.html (accessed on 22 January 2019).
59. Wulder, M.A.; White, J.C.; Loveland, T.R.; Woodcock, C.E.; Belward, A.S.; Cohen, W.B.; Fosnight, E.A.; Shaw, J.; Masek, J.G.; Roy, D.P. The global Landsat archive: Status, consolidation, and direction. *Remote Sens. Environ.* **2016**, *185*, 271–283. [[CrossRef](#)]
60. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P. Sentinel-2: ESA’s optical high-resolution mission for GMES operational services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [[CrossRef](#)]
61. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. [[CrossRef](#)]
62. Fongaro, C.; Demattê, J.; Rizzo, R.; Lucas Safanelli, J.; Mendes, W.; Dotto, A.; Vicente, L.; Franceschini, M.; Ustin, S. Improvement of clay and sand quantification based on a novel approach with a focus on multispectral satellite images. *Remote Sens.* **2018**, *10*, 1555. [[CrossRef](#)]
63. Poggio, L.; Lassaue, A.; Gimona, A. Modeling the extent of northern peat soil and its uncertainty with Sentinel: Scotland as example of highly cloudy region. *Geoderma* **2019**, *346*, 63–74. [[CrossRef](#)]
64. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]
65. Kursa, M.B.; Rudnicki, W.R. Feature selection with the Boruta package. *J. Stat. Softw.* **2010**, *36*, 1–13. [[CrossRef](#)]
66. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
67. Keskin, H.; Grunwald, S.; Harris, W.G. Digital mapping of soil carbon fractions with machine learning. *Geoderma* **2019**, *339*, 40–58. [[CrossRef](#)]
68. Wang, R. Significantly improving the prediction of molecular atomization energies by an ensemble of machine learning algorithms and rescanning input space: A stacked generalization approach. *J. Phys. Chem. C* **2018**, *122*, 8868–8873. [[CrossRef](#)]
69. Quinlan, J.R. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Tasmania, Australia, 16–18 November 1992; pp. 343–348.
70. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: Berlin, Germany, 2013; Volume 26.
71. Minasny, B.; McBratney, A.B. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chem. Intell. Lab. Syst.* **2008**, *94*, 72–79. [[CrossRef](#)]
72. Kuhn, M.; Weston, S.; Keefer, C.; Kuhn, M.M. Package ‘Cubist’. Available online: <https://cran.r-project.org/web/packages/Cubist/index.html> (accessed on 22 January 2019).
73. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. Classification and regression trees. *Wadsworth Int. Group* **1984**, *37*, 237–251.

74. Behrens, T.; Schmidt, K.; Zhu, A.-X.; Scholten, T. The ConMap approach for terrain-based digital soil mapping. *Eur. J. Soil Sci.* **2010**, *61*, 133–143. [[CrossRef](#)]
75. Hounkpatin, K.O.; Schmidt, K.; Stumpf, F.; Forkuor, G.; Behrens, T.; Scholten, T.; Amelung, W.; Welp, G. Predicting reference soil groups using legacy data: A data pruning and Random Forest approach for tropical environment (Dano catchment, Burkina Faso). *Sci. Rep.* **2018**, *8*, 9959. [[CrossRef](#)] [[PubMed](#)]
76. Wiesmeier, M.; Barthold, F.; Blank, B.; Kögel-Knabner, I. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil* **2011**, *340*, 7–24. [[CrossRef](#)]
77. Peters, J.; De Baets, B.; Verhoest, N.E.; Samson, R.; Degroeve, S.; De Becker, P.; Huybrechts, W. Random forests as a tool for ecohydrological distribution modeling. *Ecol. Model.* **2007**, *207*, 304–318. [[CrossRef](#)]
78. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [[CrossRef](#)]
79. Chen, T.; He, T.; Benesty, M.; Khotilovich, V.; Tang, Y. Xgboost: Extreme Gradient Boosting. R Package Version 0.4-2. Available online: <https://github.com/dmlc/xgboost> (accessed on 22 January 2019).
80. Venables, B.; Ripley, B. VR: Bundle of MASS, class, nnet, spatial. R package version 7.2-42. Available online: <http://CRAN.R-project.org/package=VR> (accessed on 22 January 2019).
81. Ripley, B.D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, UK, 2007.
82. Hagan, M.T.; Menhaj, M.B. Training feedforward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **1994**, *5*, 989–993. [[CrossRef](#)]
83. Meyer, H.; Kühnlein, M.; Appelhans, T.; Nauss, T. Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmos. Res.* **2016**, *169*, 424–433. [[CrossRef](#)]
84. Baker, L.; Ellison, D. Optimisation of pedotransfer functions using an artificial neural network ensemble method. *Geoderma* **2008**, *144*, 212–224. [[CrossRef](#)]
85. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **2008**, *28*, 1–26. [[CrossRef](#)]
86. Hinton, G.E.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)] [[PubMed](#)]
87. Shen, C. A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resour. Res.* **2018**, *54*, 8558–8593. [[CrossRef](#)]
88. Suthaharan, S. Big data analytics: Machine learning and Bayesian learning perspectives—What is done? What is not? *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1283. [[CrossRef](#)]
89. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
90. Zhai, J.; Zang, L.; Zhou, Z. Ensemble dropout extreme learning machine via fuzzy integral for data classification. *Neurocomputing* **2018**, *275*, 1043–1052. [[CrossRef](#)]
91. Candel, A.; Parmar, V.; LeDell, E.; Arora, A. *Deep Learning with H2O*; H2O. AI Inc.: Mountain View, CA, USA, 2016.
92. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2008**, *33*, 1–22. [[CrossRef](#)]
93. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
94. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines. *Adv. Neural Inf. Process. Syst.* **1997**, *28*, 779–784.
95. Dimitriadou, E.; Hornik, K.; Leisch, F.; Meyer, D.; Weingessel, A.; Leisch, M.F. The e1071 Package. Misc Functions of Department of Statistics (e1071), TU Wien, Vienna, Austria. *R Package Version 1.5-18*. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 22 January 2019).
96. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. mlr: Machine learning in R. *J. Mach. Learn. Res.* **2016**, *17*, 5938–5942.
97. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [[CrossRef](#)]
98. Khaledian, Y.; Miller, B.A. Selecting appropriate machine learning methods for digital soil mapping. *Appl. Math. Model.* **2019**, *81*, 401–418. [[CrossRef](#)]
99. Nawar, S.; Mouazen, A.M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena* **2017**, *151*, 118–129. [[CrossRef](#)]

100. Jafari, M.; Chahouki, M.Z.; Tavili, A.; Azarnivand, H.; Amiri, G.Z. Effective environmental factors in the distribution of vegetation types in Poshtkouh rangelands of Yazd Province (Iran). *J. Arid Environ.* **2004**, *56*, 627–641. [[CrossRef](#)]
101. Goebes, P.; Schmidt, K.; Seitz, S.; Both, S.; Bruelheide, H.; Erfmeier, A.; Scholten, T.; Kühn, P. The strength of soil-plant interactions under forest is related to a Critical Soil Depth. *Sci. Rep.* **2019**, *9*, 8635. [[CrossRef](#)]
102. Wang, B.; Waters, C.; Orgill, S.; Gray, J.; Cowie, A.; Clark, A.; Li Liu, D. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. *Sci. Total Environ.* **2018**, *630*, 367–378. [[CrossRef](#)]
103. Vaudour, E.; Gomez, C.; Fouad, Y.; Lagacherie, P. Sentinel-2 image capacities to predict common topsoil properties of temperate and Mediterranean agroecosystems. *Remote Sens. Environ.* **2019**, *223*, 21–33. [[CrossRef](#)]
104. Tziachris, P.; Aschonitis, V.; Chatzistathis, T.; Papadopoulou, M. Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena* **2019**, *174*, 206–216. [[CrossRef](#)]
105. Hengl, T.; Leenaars, J.G.; Shepherd, K.D.; Walsh, M.G.; Heuvelink, G.B.; Mamo, T.; Tilahun, H.; Berkhout, E.; Cooper, M.; Fegraus, E. Soil nutrient maps of Sub-Saharan Africa: Assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* **2017**, *109*, 77–102. [[CrossRef](#)]
106. Ramcharan, A.; Hengl, T.; Nauman, T.; Brungard, C.; Waltman, S.; Wills, S.; Thompson, J. Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Sci. Soc. Am. J.* **2018**, *82*, 186–201. [[CrossRef](#)]
107. Chen, S.; Liang, Z.; Webster, R.; Zhang, G.; Zhou, Y.; Teng, H.; Hu, B.; Arrouays, D.; Shi, Z. A high-resolution map of soil pH in China made by hybrid modeling of sparse soil data and environmental covariates and its implications for pollution. *Sci. Total Environ.* **2019**, *655*, 273–283. [[CrossRef](#)] [[PubMed](#)]
108. Hengl, T.; de Jesus, J.M.; Heuvelink, G.B.; Gonzalez, M.R.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [[CrossRef](#)] [[PubMed](#)]
109. Shangguan, W.; Hengl, T.; de Jesus, J.M.; Yuan, H.; Dai, Y. Mapping the global depth to bedrock for land surface modeling. *J. Adv. Model. Earth Syst.* **2017**, *9*, 65–88. [[CrossRef](#)]
110. Saeedimoghaddam, M.; Stepinski, T.F. Automatic extraction of road intersection points from USGS historical map series using deep convolutional neural networks. *Int. J. Geogr. Inf. Sci.* **2019**, 1–22. [[CrossRef](#)]
111. Padarian, J.; Minasny, B.; McBratney, A.B. Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional* **2019**, *16*, e00198. [[CrossRef](#)]
112. Somarathna, P.D.; Minasny, B.; Malone, B.P. More data or a better model? Figuring out what matters most for the spatial prediction of soil carbon. *Soil Sci. Soc. Am. J.* **2017**, *81*, 1413–1426. [[CrossRef](#)]
113. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Lean Publishing: Victoria, BC, Canada, 2018.
114. Elith, J.; Ferrier, S.; Huettmann, F.; Leathwick, J. The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecol. Model.* **2005**, *186*, 280–289. [[CrossRef](#)]
115. Laub, M.; Blagodatsky, S.; Lang, R.; Yang, X.; Cadisch, G. A mixed model for landscape soil organic carbon prediction across continuous profile depth in the mountainous subtropics. *Geoderma* **2018**, *330*, 177–192. [[CrossRef](#)]
116. Mishra, U.; Lal, R.; Slater, B.; Calhoun, F.; Liu, D.; Van Meirvenne, M. Predicting soil organic carbon stock using profile depth distribution functions and ordinary kriging. *Soil Sci. Soc. Am. J.* **2009**, *73*, 614–621. [[CrossRef](#)]
117. Vasques, G.; Grunwald, S.; Comerford, N.; Sickman, J. Regional modeling of soil carbon at multiple depths within a subtropical watershed. *Geoderma* **2010**, *156*, 326–336. [[CrossRef](#)]

