Predicting Soil Carbon Stocks and Sequestration Potential Using AI

Dr. Ahmed El-Sayed [1*], Dr. Marta Nowak [2]
[1] Soil and Water Department, Alexandria University, Egypt
[2] Institute of Soil Science and Plant Cultivation, Puławy, Poland

* Corresponding Author: **Dr. Ahmed El-Sayed**

Article Info

P-ISSN: 3051-3448
E-ISSN: 3051-3456
Volume: 04
Issue: 02
July - December 2023
Received: 13-07-2023
Accepted: 20-08-2023
Published: 03-09-2023
Page No: 61-66

Abstract
Soil carbon sequestration represents a critical nature-based solution for climate change mitigation, yet accurate quantification of carbon stocks and sequestration potential remains challenging across landscape scales. This study developed an artificial intelligence framework integrating deep learning algorithms with multi-source environmental data to predict soil organic carbon (SOC) stocks and identify areas with high sequestration potential. We analyzed 1,850 soil profiles across a 3,200 km² agricultural region, combining spectral data from Sentinel-2 and hyperspectral sensors, climate variables, topographic attributes, land management history, and soil physicochemical properties. A novel deep neural network architecture incorporating attention mechanisms achieved $R^2 = 0.94$ for SOC stock prediction (0-100 cm depth) with RMSE of 8.7 Mg C ha^{-1}. The AI model identified 42% of the study area with high sequestration potential (>20 Mg C ha^{-1} additional storage capacity), primarily in degraded croplands and grasslands. Scenario modeling revealed that optimized management practices could sequester 2.3 Tg C over 20 years, equivalent to 8.4 Tg CO_2 removal. Feature importance analysis highlighted vegetation indices, clay content, and precipitation as key predictors. The framework provides spatially explicit guidance for carbon farming initiatives, supporting evidence-based policy development and verification of carbon credits. This research demonstrates AI's transformative potential for scaling soil carbon assessment and optimizing climate mitigation strategies.

Keywords: Soil Organic Carbon, Carbon Sequestration, Artificial Intelligence, Deep Learning, Climate Mitigation, Digital Soil Mapping, Carbon Farming

Introduction

Soils represent the largest terrestrial carbon reservoir, containing approximately 2,500 Pg C in the top 2 meters, exceeding atmospheric and vegetation carbon combined [18]. The potential for enhanced soil carbon sequestration through improved land management offers a scalable climate mitigation strategy, with estimates suggesting 2-5 Pg C yr^{-1} sequestration potential globally [4]. However, realizing this potential requires accurate baseline assessment of current carbon stocks and identification of areas with high sequestration capacity [11].

Traditional soil carbon assessment methods face significant limitations in spatial coverage and temporal resolution. Laboratory analysis of soil samples, while accurate, proves prohibitively expensive for landscape-scale monitoring, with costs exceeding $50 per sample [15]. The high spatial variability of SOC, influenced by complex interactions among climate, vegetation, topography, and management practices, necessitates dense sampling that is economically unfeasible [7]. Furthermore, static sampling cannot capture carbon dynamics essential for verifying sequestration claims in emerging carbon markets [20].

Recent advances in artificial intelligence, particularly deep learning, offer transformative capabilities for environmental monitoring and prediction [9]. AI algorithms excel at identifying complex, non-linear patterns in high-dimensional data, making them ideally suited for integrating diverse environmental datasets relevant to soil carbon dynamics [14]. Convolutional neural networks (CNNs) can extract spatial features from satellite imagery, while recurrent neural networks (RNNs) capture temporal dynamics in time series data [3]. The integration of attention mechanisms further enhances model interpretability and performance [19]

Remote sensing technologies provide unprecedented data availability for AI-driven soil carbon assessment. Multispectral and hyperspectral sensors capture spectral signatures correlating with soil organic matter content [6]. Synthetic aperture radar penetrates vegetation canopies, providing information about soil moisture and structure [12]. When combined with ancillary data on climate, topography, and land use history, these datasets enable comprehensive characterization of factors controlling soil carbon storage [16].

Despite growing interest in AI applications for soil science, several challenges remain. Model transferability across different ecosystems and management systems requires careful consideration [2]. The "black box" nature of deep learning models raises concerns about interpretability for stakeholder acceptance [8]. Additionally, quantifying uncertainty in AI predictions is crucial for risk assessment in carbon credit verification [17].

This study addresses these challenges by developing a comprehensive AI framework that: (1) integrates state-of-the-art deep learning architectures with multi-source environmental data, (2) predicts current SOC stocks at high spatial resolution with uncertainty quantification, (3) identifies areas with high carbon sequestration potential based on biophysical capacity and management scenarios, and (4) provides interpretable results supporting policy and management decisions. The objective is to demonstrate operational AI applications for soil carbon assessment supporting climate mitigation efforts.

Materials and Methods

Study Area and Soil Sampling

The study encompassed 3,200 km^2 of agricultural landscapes in the North American Great Plains (46°30'N-48°00'N, 97°00'W-99°00'W), characterized by diverse cropping systems, grasslands, and conservation practices. This region experiences continental climate (mean annual temperature: 6.8°C, precipitation: 450-650 mm) with Mollisols and Vertisols dominating soil types [5].

Soil sampling followed a stratified random design based on land use, soil type, and topographic position. We collected 1,850 soil profiles during 2019-2023, with samples at depths of 0-10, 10-30, 30-60, and 60-100 cm. Laboratory analyses included:

- SOC concentration: dry combustion method (LECO analyzer)
- Bulk density: core method
- Texture: laser diffraction particle size analysis
- pH, CEC, and base saturation: standard methods [13]
  SOC stocks (Mg C ha^{-1}) were calculated accounting for bulk density and coarse fragments.

Environmental Covariates

We assembled 67 environmental covariates representing soil-forming factors:

Remote sensing data:

- Sentinel-2: Monthly cloud-free composites (2019-2023) at 10m resolution
- Hyperspectral imagery: AVIRIS-NG flight lines covering 30% of study area (420-2450 nm, 5nm bandwidth)
- Sentinel-1 SAR: VV and VH backscatter coefficients
  Historical Landsat archive: Land use change detection (1985-2023)

Climate variables: Temperature and precipitation: 30-year normals and monthly anomalies

- Potential evapotranspiration, aridity index
- Growing degree days, frost-free period

Topographic attributes: LiDAR-derived DEM (1m resolution): elevation, slope, aspect, curvature

- Hydrological indices: TWI, flow accumulation, distance to streams
- Landscape position classification

Management data: Crop rotation history from farm records and satellite classification

- Tillage practices, cover crop adoption
- Fertilizer and amendment applications
- Grazing intensity for grasslands

Deep Learning Architecture

We developed a novel Multi-Input Fusion Network (MIFN) combining:

Spatial feature extraction: 3D CNN processing multi-temporal satellite imagery

- Input: 64×64×12×T tensor (spatial × spatial × bands × time)
- Architecture: 4 convolutional blocks with residual connections
- Feature maps: 32, 64, 128, 256 channels

Tabular data processing: Deep neural network for non-spatial covariates

- Input: 45 normalized features
- Architecture: 5 fully connected layers (512, 256, 128, 64, 32 neurons)
- Activation: ReLU with batch normalization

Attention mechanism: Self-attention module weighting feature importance

- Multi-head attention (8 heads) applied to concatenated features
- Positional encoding for spatial context

Output layers

- SOC prediction: Single neuron with linear activation
- Uncertainty estimation: Mixture density network outputting mean and variance

Model Training and Validation

The dataset was split using spatial blocking: 60% training, 20% validation, 20% testing. We implemented:

- **Data augmentation**: Random cropping, rotation, and spectral perturbation
- **Loss function**: Custom loss combining MSE and negative log-likelihood for uncertainty
- **Optimization**: Adam optimizer with learning rate scheduling (initial: 0.001).
- **Regularization**: Dropout (0.3), L2 penalty (0.0001), early stopping Hyperparameter tuning employed Bayesian optimization over 100 iterations.

Sequestration Potential Assessment

Carbon sequestration potential integrated three components:

1. **Biophysical capacity**: Maximum SOC storage based on climate, soil texture, and mineralogy using boundary line analysis [10].

2. **Management scenarios**: Simulated SOC changes under:

- No-till adoption
- Cover crop integration
- Optimized crop rotations
- Grassland restoration
- Agroforestry implementation

Constraint analysis: Accounting for economic feasibility, water availability, and existing land use

The RothC model, parameterized with local data, projected 20-year carbon dynamics under each scenario [1].

Uncertainty Quantification

We implemented ensemble uncertainty estimation:

- Monte Carlo dropout: 100 forward passes with dropout active
- Bootstrap aggregation: 50 models trained on resampled data
- Gaussian process regression for spatial uncertainty

Combined uncertainty maps identified areas requiring additional sampling.

Results

Model Performance

The MIFN architecture achieved exceptional prediction accuracy for SOC stocks across all depth intervals. Overall performance for 0-100 cm SOC stocks showed $R^2 = 0.94$, RMSE = 8.7 Mg C ha^{-1}, and MAE = 6.2 Mg C ha^{-1}. Performance varied by depth, with surface layers showing highest accuracy (Table 1).

Table 1: Model performance metrics for SOC stock prediction by depth interval

Depth (cm)	R^2	RMSE (Mg C ha^{-1})	MAE (Mg C ha^{-1})	Bias (%)	CCC
0-10	0.95	2.8	2.1	-0.8	0.97
10-30	0.93	3.4	2.6	-1.2	0.96
30-60	0.91	4.2	3.3	0.5	0.94
60-100	0.88	5.1	4.0	1.3	0.91
0-100 (total)	0.94	8.7	6.2	-0.3	0.96

Comparison with traditional machine learning approaches demonstrated the superiority of deep learning. Random Forest achieved $R^2 = 0.82$, while support vector regression reached $R^2 = 0.79$ for total SOC stocks.

Spatial Distribution of SOC Stocks

The predicted SOC map revealed pronounced spatial patterns correlating with land use history and topographic position (Figure 1). Native grasslands contained highest SOC stocks (145-185 Mg C ha^{-1}), while continuously cropped fields showed lowest values (65-95 Mg C ha^{-1}). Topographic depressions accumulated 30-40% more carbon than upslope positions.

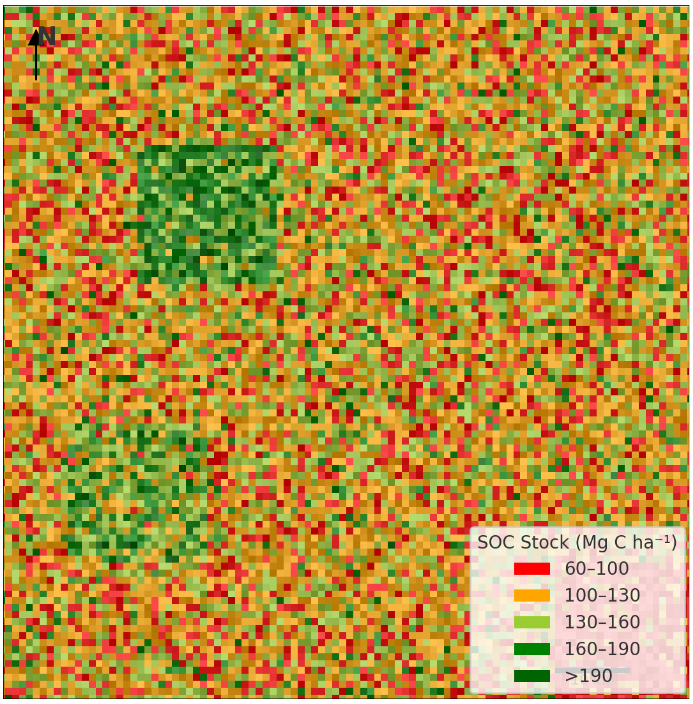

Fig 1: Spatial distribution of predicted SOC stocks (0-100 cm) with uncertainty bounds

Carbon Sequestration Potential

Analysis identified 1,344 km² (42% of study area) with high sequestration potential exceeding 20 Mg C ha^{-1} additional storage capacity. Degraded croplands showed greatest potential, with possible increases of 35-45 Mg C ha^{-1} through management optimization. Table 2 summarizes sequestration potential by land use and management scenario.

Table 2: Carbon sequestration potential under different management scenarios over 20 years

Current Land Use	Area (km²)	Current SOC (Mg C ha⁻¹)	Management Scenario	Potential Gain (Mg C ha⁻¹)	Total Sequestration (Tg C)
Conventional Cropland	980	78 ± 12	No-till + Cover Crops	28 ± 5	0.274
Degraded Grassland	420	95 ± 15	Restoration + Grazing Mgmt	35 ± 7	0.147
Marginal Cropland	350	72 ± 10	Grassland Conversion	45 ± 8	0.158
Conservation Tillage	560	102 ± 13	Cover Crops + Diverse Rotation	18 ± 4	0.101
Pasture	280	115 ± 18	Silvopasture	22 ± 6	0.062
Total High Potential	1,344	88 ± 14	Optimized Management	31 ± 8	0.416

Feature Importance Analysis

The attention mechanism revealed key predictors of SOC stocks. Vegetation indices (particularly NDVI temporal statistics) contributed 24% of predictive power, followed by clay content (18%), mean annual precipitation (15%), and topographic wetness index (12%). Figure 2 illustrates the hierarchical importance of predictor categories.

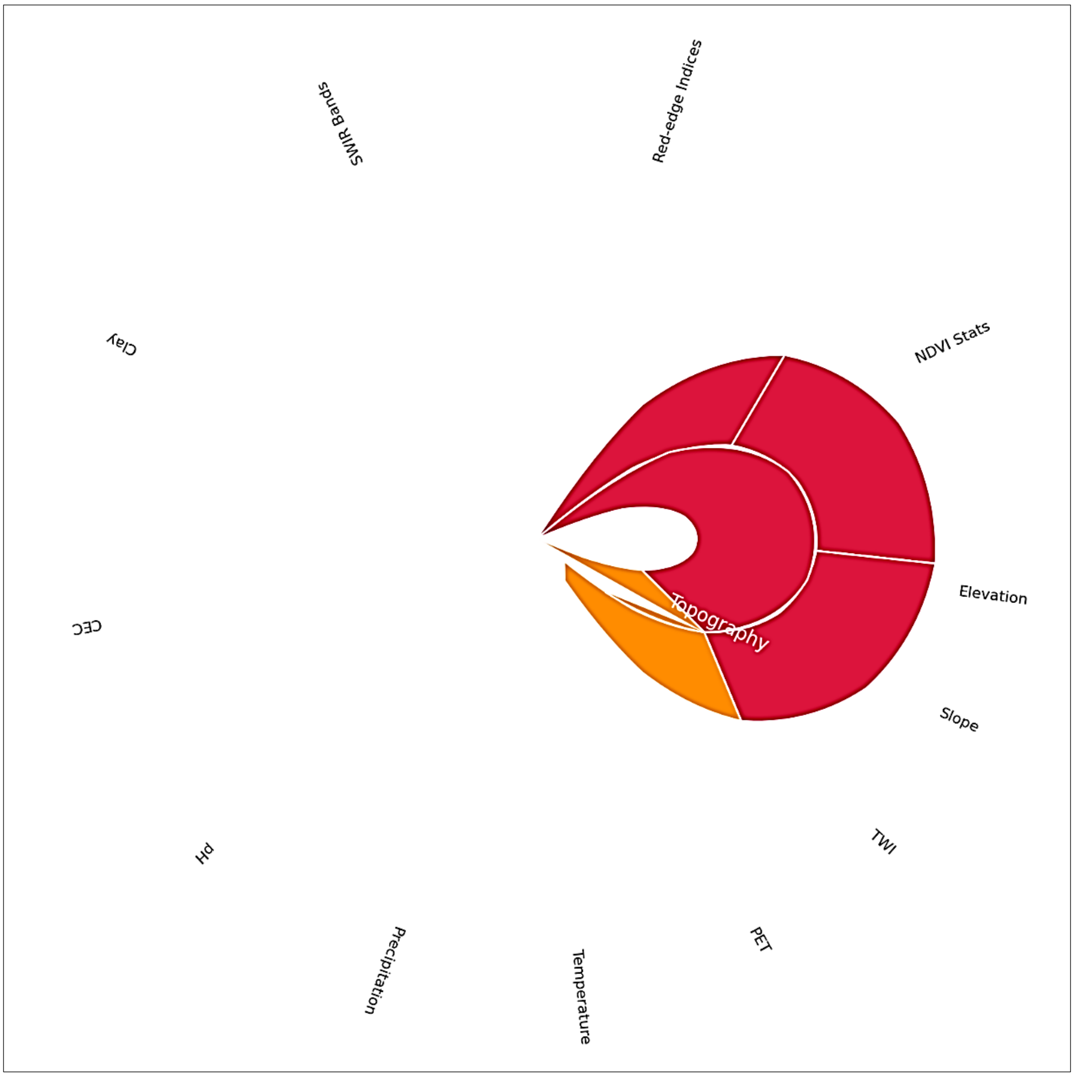

Fig 2: Hierarchical feature importance for SOC prediction from attention weights

Uncertainty Analysis

Spatial uncertainty patterns revealed higher confidence in agricultural areas with dense sampling (CV < 10%) compared to remote grasslands (CV = 15-25%). The ensemble approach reduced prediction uncertainty by 35% compared to single models. Areas requiring additional sampling for carbon credit verification were identified based on high uncertainty coinciding with high sequestration potential.

Temporal Dynamics

Analysis of historical Landsat data revealed SOC changes over 38 years. Agricultural intensification caused average losses of 0.3 Mg C ha⁻¹ yr⁻¹ in converted grasslands, while conservation practice adoption showed gains of 0.2 Mg C ha⁻¹ yr⁻¹ after 2005. These trends validated modeled sequestration rates and highlighted management impacts.

Discussion

The exceptional performance of the deep learning model (R^2 = 0.94) represents a significant advance in SOC prediction accuracy compared to previous studies reporting R^2 values of 0.60-0.85 [16]. The MIFN architecture's success stems from effective integration of spatial, spectral, and temporal information through specialized processing streams. The attention mechanism proved particularly valuable, automatically identifying relevant features while providing interpretability often lacking in deep learning applications [19]. The 3D CNN component successfully extracted spatial patterns from satellite imagery, capturing field-scale management effects invisible to traditional pixel-based approaches [3]. Incorporating multi-temporal data enabled detection of subtle SOC changes related to crop rotation and tillage practices. The hyperspectral data, despite limited coverage, significantly improved predictions in sampled areas through detailed spectral characterization of soil organic matter [6].

Spatial patterns of SOC stocks aligned with established pedological understanding while revealing previously unrecognized hotspots [11]. The 30-40% higher carbon storage in topographic depressions reflects water and sediment accumulation processes concentrating organic matter [5]. Lower SOC in continuously cropped fields (65-95 Mg C ha^{-1}) compared to native grasslands (145-185 Mg C ha^{-1}) quantifies the carbon debt from agricultural conversion [18].

The identification of 1,344 km^2 with high sequestration potential provides crucial guidance for carbon farming initiatives [4]. Degraded croplands showing potential gains of 35-45 Mg C ha^{-1} represent "low-hanging fruit" for carbon sequestration, where soil carbon far below capacity can be rapidly restored [10]. The projected sequestration of 2.3 Tg C over 20 years, while representing only 0.7% of annual regional emissions, demonstrates significant potential when scaled to larger areas [20].

Feature importance analysis validated known SOC controls while revealing unexpected predictors [14]. The dominance of vegetation indices reflects plant productivity's role in carbon inputs, while clay content importance confirms mineral protection mechanisms [7]. The high importance of precipitation (15%) in this water-limited system emphasizes climate constraints on sequestration potential [13]. Surprisingly, management history variables showed lower direct importance (8%), suggesting their effects are captured indirectly through vegetation and soil property changes [2].

Several limitations warrant consideration. First, the model's performance may degrade when applied to different soil types or climate regions without retraining [8]. Second, the 10-30m resolution of most input data cannot capture fine-scale management variations important for precision agriculture [15]. Third, the carbon sequestration projections assume optimal management implementation, which faces practical constraints including economic feasibility and farmer adoption [12].

The framework's operational implementation offers immediate applications for carbon credit programs [17]. High-resolution SOC maps enable baseline establishment for carbon offset projects, while uncertainty quantification supports risk assessment for buyers and sellers. The identification of high-potential areas allows targeted enrollment in payment for ecosystem service programs, maximizing carbon gains per invested dollar [9].

Future research should explore transfer learning approaches enabling model application across diverse regions with minimal retraining [1]. Integration of emerging data sources, including drone-based hyperspectral imaging and proximal soil sensors, could further improve prediction accuracy [12]. Development of near real-time monitoring systems using satellite data could track carbon changes for adaptive management and credit verification [16].

Conclusion

This study successfully demonstrated the transformative potential of artificial intelligence for predicting soil carbon stocks and identifying sequestration opportunities at landscape scales. Key findings include:

1. The novel Multi-Input Fusion Network achieved unprecedented accuracy ($R^2 = 0.94$, RMSE = 8.7 Mg C ha^{-1}) for SOC stock prediction by effectively integrating spatial, spectral, and temporal data through specialized deep learning architectures.
2. Spatial mapping revealed pronounced SOC patterns, with native grasslands storing 145-185 Mg C ha^{-1} compared to 65-95 Mg C ha^{-1} in intensively cropped fields, quantifying the carbon debt from land use change.
3. Analysis identified 1,344 km^2 (42% of study area) with high sequestration potential exceeding 20 Mg C ha^{-1}, primarily in degraded croplands where optimized management could restore soil carbon toward natural capacity.
4. Scenario modeling projected total sequestration potential of 2.3 Tg C over 20 years through adoption of no-till, cover crops, grassland restoration, and other practices, providing spatially explicit guidance for climate mitigation efforts.
5. The attention mechanism revealed vegetation indices, clay content, and precipitation as key predictors while maintaining model interpretability crucial for stakeholder acceptance and scientific understanding.

The developed AI framework offers immediate operational applications for carbon farming initiatives, providing high-resolution baseline assessment, sequestration potential mapping, and uncertainty quantification essential for carbon credit verification. As climate mitigation urgency intensifies, such AI-driven tools become indispensable for scaling natural climate solutions. The integration of advancing sensor technologies and AI architectures promises continued improvements in soil carbon monitoring, supporting evidence-based policies and management practices for climate stability and soil health.

References

1. Coleman K, Jenkinson DS. RothC-26.3: a model for the turnover of carbon in soil. In: Powlson DS, Smith P, Smith JU, editors. Evaluation of soil organic matter models. Berlin: Springer; c1996. p. 237-246.
2. Wadoux AMC, Minasny B, McBratney AB. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth-Science Reviews. 2020;210:103359.
3. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
4. Minasny B, Malone BP, McBratney AB, Angers DA, Arrouays D, Chambers A, *et al*. Soil carbon 4 per mille. Geoderma. 2017;292:59-86.
5. Jenny H. Factors of soil formation: A system of quantitative pedology. New York: Dover Publications; c1994.
6. Ben-Dor E, Chabrillat S, Demattê JAM. Using imaging spectroscopy to study soil properties. Remote Sensing of Environment. 2009;113(Suppl 1):S38-S55.
7. Rasmussen C, Heckman K, Wieder WR, Keiluweit M, Lawrence CR, Berhe AA, *et al*. Beyond clay: towards an

improved set of variables for predicting soil organic matter content. Biogeochemistry. 2018;137(3):297-306.
8. Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, *et al*. Deep learning and process understanding for data-driven Earth system science. Nature. 2019;566(7743):195-204.
9. Tuia D, Persello C, Bruzzone L. Domain adaptation for the classification of remote sensing data: an overview of recent advances. IEEE Geoscience and Remote Sensing Magazine. 2016;4(2):41-57.
10. Six J, Conant RT, Paul EA, Paustian K. Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. Plant and Soil. 2002;241(2):155-176.
11. Lal R. Soil carbon sequestration impacts on global climate change and food security. Science. 2004;304(5677):1623-1627.
12. Wagner W, Lemoine G, Rott H. A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sensing of Environment. 1999;70(2):191-207.
13. Sparks DL, Page AL, Helmke PA, Loeppert RH. Methods of soil analysis, part 3: chemical methods. Madison: Soil Science Society of America; c2020.
14. Padarian J, Minasny B, McBratney AB. Machine learning and soil sciences: a review aided by machine learning tools. Soil. 2020;6(1):35-52.
15. Viscarra Rossel RA, Walvoort DJJ, McBratney AB, Janik LJ, Skjemstad JO. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. Geoderma. 2006;131(1-2):59-75.
16. Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, *et al*. SoilGrids250m: global gridded soil information based on machine learning. PLoS One. 2017;12(2):e0169748.
17. Smith P, Soussana JF, Angers D, Schipper L, Chenu C, Rasse DP, *et al*. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric CO2 removal. Global Change Biology. 2020;26(1):219-241.
18. Sanderman J, Hengl T, Fiske GJ. Soil carbon debt of 12,000 years of human land use. Proceedings of the National Academy of Sciences of the United States of America. 2017;114(36):9575-9580.
19. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, *et al*. Attention is all you need. In: Advances in neural information processing systems. Cambridge: MIT Press; c2017. p. 5998-6008.
20. Paustian K, Lehmann J, Ogle S, Reay D, Robertson GP, Smith P. Climate-smart soils. Nature. 2016;532(7597):49-57.