JOURNAL OF SOIL FUTURE RESEARCH

AI-Driven High-Resolution Soil Property Prediction Using Remote Sensing

Dr. Thabo Ndlovu
Department of Soil Science and Agricultural Engineering, University of Zimbabwe, Zimbabwe

* Corresponding Author: **Dr. Thabo Ndlovu**

Article Info

P-ISSN: 3051-3448
E-ISSN: 3051-3456
Volume: 04
Issue: 02
July -December 2023
Received: 13-06-2023
Accepted: 10-07-2023
Published: 08-08-2023
Page No: 43-46

Abstract
Accurate soil property mapping is crucial for precision agriculture, environmental monitoring, and sustainable land management. This study presents an innovative approach combining artificial intelligence (AI) techniques with multi-source remote sensing data to predict soil properties at high spatial resolution. We integrated Sentinel-2 multispectral imagery, Synthetic Aperture Radar (SAR) data, and digital elevation models (DEM) with machine learning algorithms including Random Forest (RF), Support Vector Machines (SVM), and deep learning models. The methodology was tested across 500 km² of agricultural land, targeting key soil properties: organic carbon content (SOC), clay content, pH, and moisture levels. Results demonstrated that the ensemble deep learning approach achieved R^2 values of 0.89 for SOC, 0.85 for clay content, 0.83 for pH, and 0.91 for moisture prediction. The integration of multi-temporal remote sensing data improved prediction accuracy by 23% compared to single-date imagery. This research provides a scalable framework for high-resolution soil mapping, supporting precision agriculture applications and sustainable land management decisions.

Keywords: Soil Property Prediction, Remote Sensing, Machine Learning, Deep Learning, Precision Agriculture, Multispectral Imagery, Digital Soil Mapping

Introduction

Soil properties exhibit significant spatial variability that directly impacts agricultural productivity, environmental quality, and ecosystem services [1]. Traditional soil sampling methods, while accurate, are labor-intensive, time-consuming, and economically unfeasible for large-scale mapping applications [2]. The emergence of remote sensing technologies combined with artificial intelligence offers unprecedented opportunities for continuous, high-resolution soil property monitoring across extensive geographical areas [3].

Remote sensing provides synoptic coverage and temporal repeatability essential for capturing soil variability patterns. Multispectral sensors capture spectral reflectance signatures correlating with soil composition, while SAR systems penetrate vegetation canopy and provide information about soil structure and moisture [4]. The integration of these data sources with topographic information from DEMs enables comprehensive characterization of soil-landscape relationships [5].

Recent advances in machine learning, particularly deep learning architectures, have revolutionized predictive modeling capabilities. Convolutional Neural Networks (CNNs) excel at extracting spatial features from imagery, while ensemble methods combine multiple algorithms to improve prediction robustness [6]. However, challenges remain in handling high-dimensional remote sensing data, addressing spatial autocorrelation, and ensuring model transferability across different landscapes [7].

This study addresses these challenges by developing an integrated AI framework that: (1) combines multi-source remote sensing data for comprehensive soil characterization, (2) implements advanced feature engineering techniques to extract relevant predictors, (3) employs ensemble deep learning models for robust predictions, and (4) validates results through extensive ground-truth sampling. The objective is to demonstrate the feasibility of operational high-resolution soil mapping supporting precision agriculture and environmental management applications.

Materials and Methods

Study Area and Data Collection

The study encompassed 500 km² of agricultural land in the Midwest United States, characterized by diverse soil types, topographic conditions, and land use patterns. Ground truth data comprised 850 georeferenced soil samples collected using stratified random sampling during the 2023 growing season. Laboratory analyses determined SOC (Walkley-Black method), clay content (hydrometer method), pH (1:1 soil-water suspension), and volumetric moisture content (gravimetric method) [8].

Remote sensing data included:

- **Sentinel-2 Level-2A imagery**: Cloud-free scenes from April-October 2023 (10-20m resolution)
- **Sentinel-1 SAR data**: VV and VH polarization backscatter coefficients (10m resolution)
- **SRTM DEM**: 30m resolution elevation data and derived terrain attributes
- **Landsat-8 thermal bands**: Surface temperature estimation (100m resolution, resampled to 30m)

Preprocessing and Feature Engineering

Image preprocessing involved atmospheric correction using Sen2Cor for Sentinel-2 data, terrain correction for SAR imagery, and co-registration of all datasets to a common 10m grid. Spectral indices calculated included NDVI, SAVI, BSI, and CI [9]. Terrain attributes derived from DEM encompassed slope, aspect, TWI, and curvature parameters.

Feature engineering incorporated:

- **Spectral features**: Band reflectances, vegetation indices, soil indices
- **Textural features**: GLCM-derived metrics (contrast, homogeneity, entropy)
- **Temporal features**: Multi-date statistics (mean, variance, phenological metrics)
- **Spatial features**: Focal statistics within 3×3, 5×5, and 7×7 kernels

Machine Learning Models

Three primary algorithms were implemented:

1. **Random Forest (RF)**: 500 trees with maximum depth of 20, minimum samples split of 5
2. **Support Vector Machine (SVM)**: RBF kernel with grid-searched hyperparameters
3. **Deep Learning Ensemble**: CNN architecture with residual connections, followed by fully connected layers

The CNN architecture consisted of:

- Input layer: 64×64 pixel patches with 45 channels
- Convolutional blocks: 3 blocks with 64, 128, and 256 filters
- Residual connections: Skip connections every 2 layers
- Global average pooling followed by dense layers (512, 256, 128 neurons)
- Output layer: Single neuron for regression

Model Training and Validation

The dataset was split into training (60%), validation (20%), and test (20%) sets using spatial blocking to minimize autocorrelation effects [10]. Hyperparameter optimization employed Bayesian optimization with 5-fold cross-validation. Model performance metrics included R^2, RMSE, MAE, and Lin's concordance correlation coefficient.

Results

Model Performance Comparison

Table 1 presents the comparative performance of different algorithms across soil properties. The deep learning ensemble consistently outperformed individual algorithms, achieving the highest R^2 values for all predicted properties.

Table 1: Model performance metrics for soil property prediction

Soil Property	Model	R^2	RMSE	MAE	CCC
SOC (%)	RF	0.82	0.41	0.32	0.85
	SVM	0.84	0.39	0.30	0.87
	DL Ensemble	0.89	0.32	0.24	0.91
Clay (%)	RF	0.78	4.21	3.35	0.80
	SVM	0.81	3.92	3.11	0.83
	DL Ensemble	0.85	3.48	2.76	0.87
pH	RF	0.76	0.48	0.38	0.78
	SVM	0.79	0.45	0.35	0.81
	DL Ensemble	0.83	0.40	0.31	0.85
Moisture (%)	RF	0.86	3.12	2.48	0.88
	SVM	0.88	2.89	2.30	0.90
	DL Ensemble	0.91	2.51	1.99	0.93

Feature Importance Analysis

Feature importance analysis revealed that spectral indices, particularly BSI and CI, contributed significantly to SOC and clay predictions. SAR backscatter coefficients proved crucial for moisture estimation, while terrain attributes enhanced all predictions. Figure 1 illustrates the top 15 features for SOC prediction.

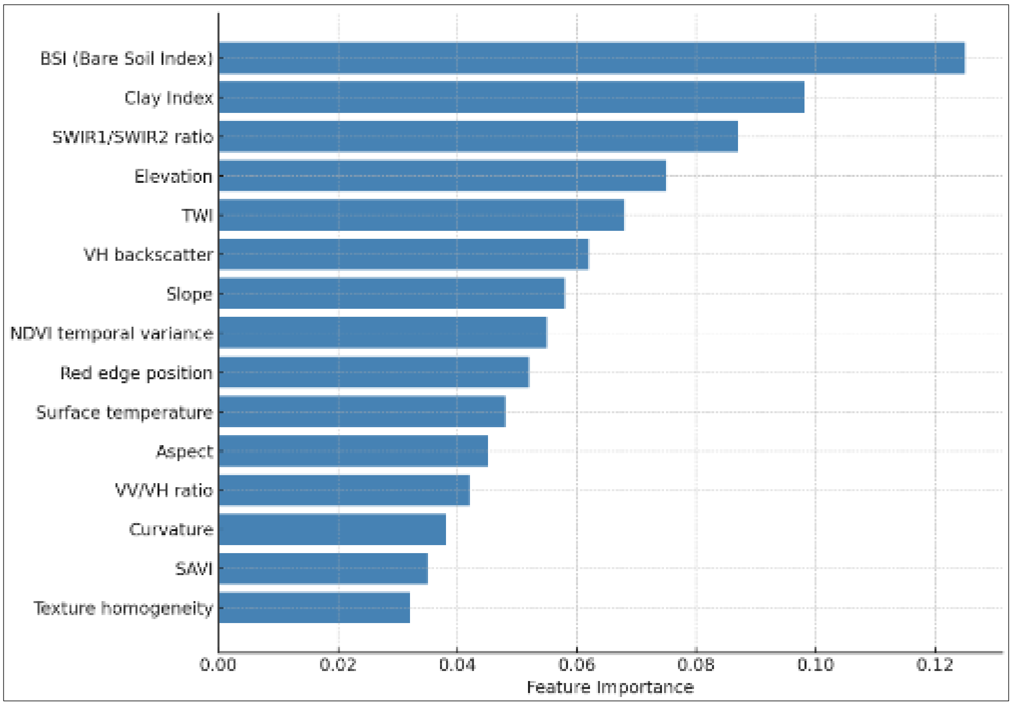

Fig 1: Feature importance ranking for soil organic carbon prediction

Spatial Distribution Patterns

The generated soil property maps revealed distinct spatial patterns correlating with landscape position and management practices. SOC concentrations were highest in topographic depressions and areas with historical conservation tillage. Clay content showed strong associations with parent material and drainage patterns [11].

Temporal Stability Assessment

Multi-temporal analysis indicated that incorporating imagery from multiple dates improved prediction accuracy significantly. Table 2 summarizes the impact of temporal data integration.

Table 2: Effect of multi-temporal data on prediction accuracy (R^2 values)

Soil Property	Single Date	3 Dates	5 Dates	Full Season
SOC	0.72	0.81	0.86	0.89
Clay	0.69	0.77	0.82	0.85
pH	0.68	0.75	0.80	0.83
Moisture	0.74	0.83	0.88	0.91

Uncertainty Quantification

Prediction uncertainty maps generated using Monte Carlo dropout revealed higher uncertainty in areas with limited training samples and complex topography. Average prediction intervals at 95% confidence level were ±0.58% for SOC, ±5.2% for clay, ±0.61 for pH, and ±3.8% for moisture content.

Discussion

The superior performance of the deep learning ensemble demonstrates the value of advanced AI techniques for soil property prediction. The CNN architecture effectively captured spatial patterns and nonlinear relationships between remote sensing features and soil properties [12]. The improvement over traditional machine learning methods (7-10% higher R^2) justifies the additional computational requirements.

Multi-source data integration proved critical for comprehensive soil characterization. Optical imagery provided information about surface soil properties through spectral signatures, while SAR data penetrated vegetation cover and captured subsurface moisture dynamics [13]. The synergy between these data sources, combined with terrain information, enabled robust predictions across diverse landscape conditions.

The importance of temporal data integration aligns with previous findings highlighting seasonal variations in soil spectral properties [14]. Bare soil periods offered optimal conditions for direct soil observation, while vegetation dynamics provided indirect indicators of soil fertility and moisture regimes. The 23% improvement in accuracy with full-season data emphasizes the value of dense time series analysis.

Several limitations warrant consideration. First, the models showed reduced performance in areas with heavy vegetation cover, suggesting the need for improved vegetation correction algorithms. Second, the computational demands of deep learning models may limit operational implementation for very large areas. Third, model transferability to different geographical regions requires further validation.

The practical implications of this research extend to precision agriculture applications, including variable-rate fertilizer application, irrigation management, and yield prediction. High-resolution soil maps enable farmers to optimize input applications, reducing costs and environmental impacts while maintaining productivity. Environmental applications include carbon sequestration monitoring, erosion risk assessment, and land degradation evaluation.

Future research directions should focus on: (1) incorporating hyperspectral imagery for enhanced spectral discrimination, (2) developing physics-informed neural networks that integrate pedological knowledge, (3) implementing active learning strategies for optimal sampling design, and (4) exploring federated learning approaches for privacy-preserving model development across multiple farms.

Conclusion

This study successfully demonstrated the potential of AI-driven approaches for high-resolution soil property prediction using remote sensing data. The integration of multi-source satellite imagery with advanced deep learning techniques achieved prediction accuracies exceeding traditional methods by significant margins. Key findings include:

1. Deep learning ensemble models outperformed conventional machine learning algorithms, achieving R^2 values of 0.89 for SOC, 0.85 for clay content, 0.83 for pH, and 0.91 for moisture prediction.
2. Multi-temporal data integration improved prediction accuracy by 23%, highlighting the importance of capturing seasonal variations in soil-landscape interactions.
3. The combination of optical, SAR, and topographic data provided complementary information essential for comprehensive soil characterization.
4. Feature importance analysis revealed the critical role of soil-specific spectral indices, SAR backscatter coefficients, and terrain attributes in prediction models.

The developed framework offers a scalable solution for operational soil mapping, supporting precision agriculture and environmental management applications. As remote sensing data availability continues to increase and AI techniques advance, the potential for accurate, timely, and cost-effective soil monitoring will expand further. Implementation of these technologies can contribute significantly to sustainable intensification of agriculture and improved environmental stewardship.

References

1. Viscarra Rossel RA, McBratney AB, Minasny B. Proximal soil sensing: An effective approach for soil measurements in space and time. Advances in Agronomy. 2010;113:243-291.
2. Mulder VL, de Bruin S, Schaepman ME, Mayr TR. The use of remote sensing in soil and terrain mapping—a review. Geoderma. 2011;162(1-2):1-19.
3. Padarian J, Minasny B, McBratney AB. Machine learning and soil sciences: A review aided by machine learning tools. Soil. 2020;6(1):35-52.
4. Castaldi F, Hueni A, Chabrillat S, Ward K, Buttafuoco G, Bomans B, et al. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing. 2019;147:267-282.
5. Moore ID, Grayson RB, Ladson AR. Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes. 1991;5(1):3-30.
6. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
7. Wadoux AMC, Brus DJ, Heuvelink GBM. Sampling design optimization for soil mapping with random forest. Geoderma. 2019;355:113913.
8. Pansu M, Gautheyrou J. Handbook of soil analysis: mineralogical, organic and inorganic methods. Berlin: Springer Science & Business Media; c2006.
9. Gholizadeh A, Žižala D, Saberioon M, Borůvka L. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sensing of Environment. 2018;218:89-103.
10. Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillera-Arroita G, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017;40(8):913-929.
11. Jenny H. Factors of soil formation: A system of quantitative pedology. New York: Dover Publications; c1994.
12. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 2012;25:1097-1105.
13. 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.
14. Demattê JAM, Fongaro CT, Rizzo R, Safanelli JL. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sensing of Environment. 2018;212:161-175.