SYNERGISTIC USE OF LOW-COST NIR SCANNER AND GEOSPATIAL COVARIATES TO ENHANCE SOIL ORGANIC CARBON PREDICTIONS USING DUAL INPUT DEEP LEARNING TECHNIQUES

Ioannis Gallios (1), Nikolaos Tziolas (1)

(1) Department of Soil, Water, and Ecosystem Sciences, Institute of Food and Agricultural Sciences, University of Florida, USA

INTRODUCTION

The significance of soil health, especially in climate-smart agriculture and carbon farming, highlights the necessity for effective soil organic carbon (SOC) monitoring. Although micro-electro miniaturized systems have enhanced reflectance infrared spectroscopy, handheld sensors still lag laboratory spectrometers in predictive accuracy. Earth Observation (EO) platforms and Artificial Intelligence (AI) algorithms can indeed complement in-situ datasets to enhance predictive performance. However, an obstacle in digital soil mapping lies in effectively fusing field spectra data with environmental and topographical information gathered from satellite sensors. AI algorithms, particularly those within the realm of deep learning (DL), exhibit notable promise in their ability to directly apprehend intricate data patterns through the utilization of convolution operations [1].

This study proposes a novel methodological framework that synergistically combines VNIR data, from a handheld spectral device, with open geospatial covariates using two distinct neural networks to enhance SOC predictions.

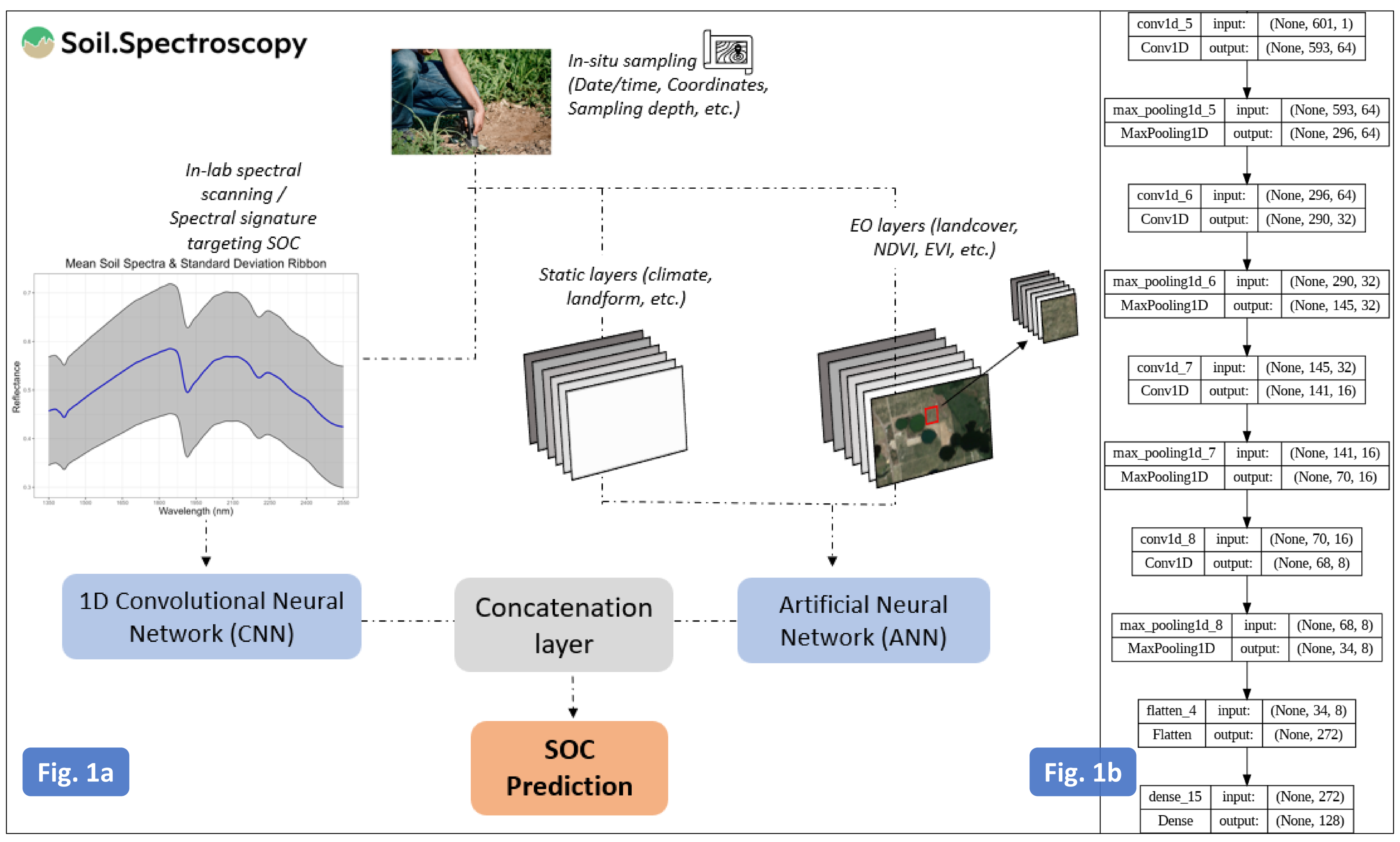

Figure 1 a) *Overall approach; illustrating how geo-covariates and NeoSpectra's spectral recordings are efficiently handled by two diverse neural networks that are concatenated to generate the SOC predictions;* ***b)*** *The architecture of the CNN was employed for the extraction of information from the spectra generated by soil samples.*

RESULTS

- XGBoost algorithm demonstrated moderate accuracy metrics, RMSE = 0.485 on the log-SOC scale.
- The use of the geo-covariates improved the overall performance of the AI models.
- Dual input approach enhances significantly predictive accuracy, **reducing the RMSE to 0.255** on the log-SOC scale (Figure 2).

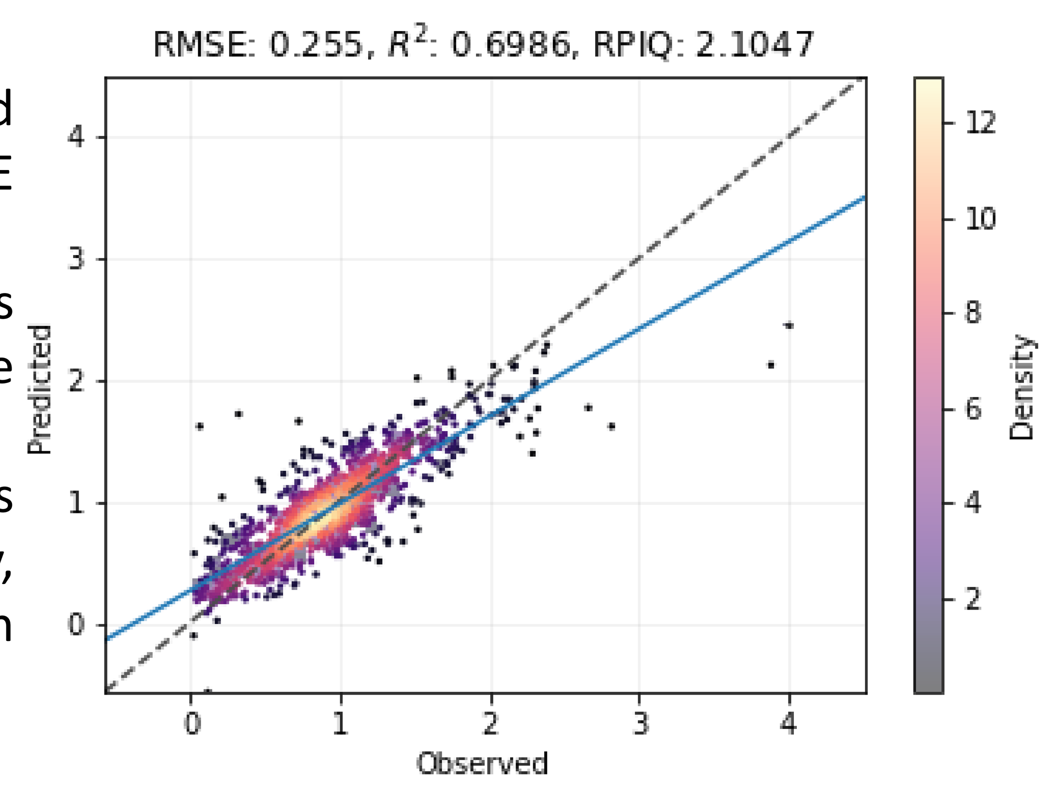

Figure 2: *Scatter plot with a heatmap overlay showcasing the relationship between observed and predicted SOC values. The x-axis, y-axis, and colormap represent the observed values, the predicted values, and the density of points respectively.*

MATERIALS & METHODS

The methodology of this work consists of three main components as depicted in Figure 1a.

Soil Spectral Data: The NeoSpectra database, accessible through the Open Soil Spectral Library, served as the primary data source for model calibration.

- A curated collection of 1202 soil samples, including SOC (dry compaction method) content was utilized; each sample came with spectra spanning from 1350 to 2550 nm.
- Independent test set derived from spectral samples across farms in Massachusetts and New York states was utilized (n=772).

Spatial Layers:

- 214 open raster spatial layers representing various factors for both the calibration and test sets.
- Information related to landform, climatic variables, vegetation indices, and land cover statistics sourced from reputable studies and datasets [2].

AI for Regression Analysis: This study introduces a dual-input DL framework that concatenates two distinct streams into a unified vector for predicting SOC.

- A convolutional neural network (CNN) for processing spectral signatures. The model includes a series of convolutional and max-pooling layers followed by dense layers for feature extraction and prediction (Fig. 1b).
- An artificial neural network (ANN) for handling the geo-covariates.
- Conventional AI algorithms, like Extreme gradient boosting (XGBoost), were compared to the proposed method for performance evaluation.

CONCLUSION

The driving force behind this research is to enhance the efficiency and accuracy of soil spectral analytics for cost-effective SOC prediction, benefiting growers engaged in carbon farming and precision agriculture. Utilizing dual-input architectures, diverse data can be efficiently managed, ensuring the interpretability of results while achieving improved accuracy compared to conventional ML models (with a 47.42% RMSE improvement).

Future research could focus on evaluating performance across additional soil parameters (e.g., pH) and integrating multispectral data from bare soil reflectance composites into geospatial layers to further enhance predictions. The study was conducted as part of the modeling challenge for the Soil Spectroscopy 4 Global Good project.

ACKNOWLEDGEMENTS

Funded by the FLA-SWF-006381 Hatch project, supported by the USDA NIFA. The NeoSpectra NIR topsoil dataset used in this work was provided by the Soil Spectroscopy for Global Good Project managed by the Woodwell Climate Research Center.

REFERENCES

[1] Karyotis, K.; Tsakiridis, N.L.; Tziolas, N.; Samarinas, N.; Kalopesa, E.; Chatzimisios, and P.; Zalidis, G. On-Site Soil Monitoring Using Photonics-Based Sensors and Historical Soil Spectral Libraries. Remote Sens. 2023, 15, 1624. https://doi.org/10.3390/rs15061624

[2] Hengl, T., 2018. Global landform and lithology class at 250 m based on the USGS global ecosystem map (1.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1464846

@SoilAILab Soil.Spectroscopy

Soil Science Artificial Intelligence Lab

SOIL, WATER, AND ECOSYSTEM SCIENCES