Enhancing Soil Organic Carbon Prediction: A Deep Learning Approach Utilizing Stacked Autoencoder Feature Extraction and Large-Scale Spectral Libraries

Mohammadmehdi Saberioon[a,b,*], Asa Gholizadeh[c], Ali Ghaznavi[d], Sabine Chabrillat[a,e], Vahid Khosravi[c]

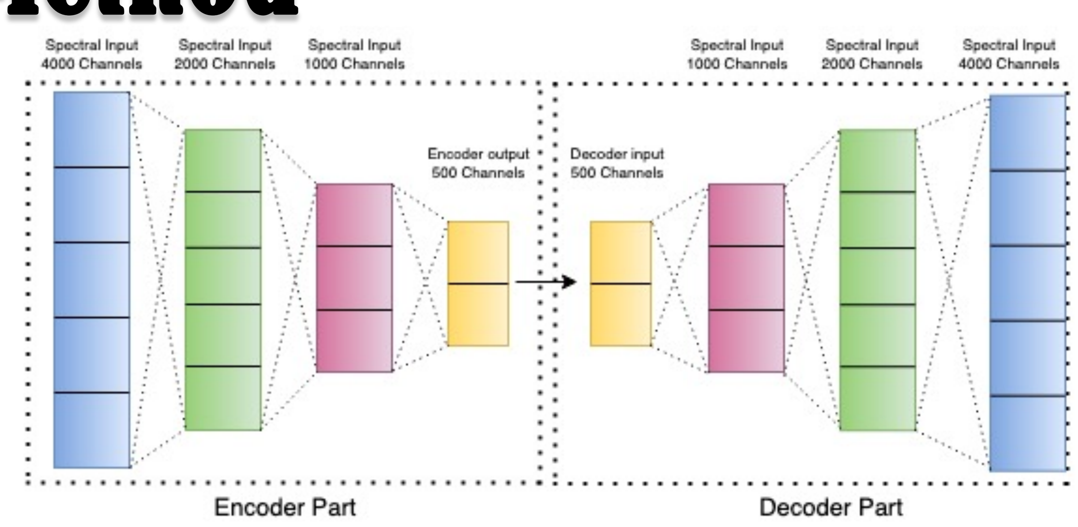
[a] Section 1.4 Remote Sensing and Geoinformatics, Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam 14473, Germany
[b]ILVO, Flanders Research Institute for Agriculture, Fisheries and Food, Technology and Food Science-Agricultural Engineering, 9820 Merelbeke, Belgium
[c]Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamycka 129, Suchdol, Prague 16500, Czech Republic
[d]Faculty of Fisheries and Protection of Waters, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, University of South Bohemia in Ceske Budejovice, Zamek 136, 373 33 Nove Hrady, Czech Republic
[e]Institute of Soil Science, Leibniz University Hannover, Herrenhauser Straße 2, 30419 Hannover, Germany

* saberioon@gfz-potsdam.de, mohammadmehdi.saberioon@ilvo.vlaanderen.be

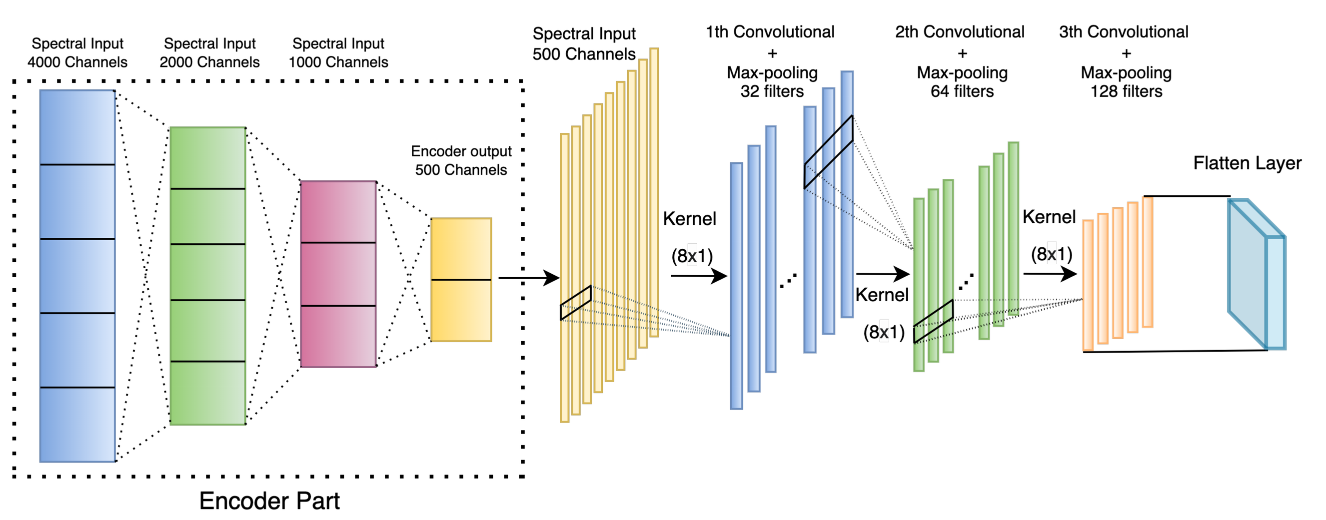
Summary

The main terrestrial carbon (C) fraction is soil organic carbon (SOC), which has a significant impact on climate change and greenhouse gases emission through the absorption and sequestration of carbon dioxide (CO_2. This has made SOC assessment very important from both economic and environmental viewpoints. The increasing number of soil spectral libraries (SSLs) on regional, continental, and global scales has brought a tremendous opportunity for quantification of SOC through the development of spectral-based prediction models. Hence, with the development of large SSLs, we need to seize the opportunity to utilize big data analytics to process the spectral data. The unique ability of deep learning (DL) techniques to leverage important features of high dimensional large-scale SSLs has made them top-demanding for more sophisticated modelling. The main objective of this study was to evaluate the capability of two different DL algorithms, i.e., one-dimensional convolutional neural network (1DCNN) and fully connected neural network (FCNN) coupled with stacked autoencoder (SAE) feature extraction for SOC prediction based on the data from the land use/cover area frame statistical survey (LUCAS) database. SAE extracted the high-level deep features from the visible--near infrared--shortwave infrared (Vis--NIR--SWIR) reflectance spectra of 11,441 soil samples, which were then considered as inputs to the 1DCNN and FCNN models for predicting the SOC content. Both SAE-DL feature-selected models yielded higher accuracy than those the DL developed on the entire spectra and a random forest (RF) model was constructed for comparison. The best prediction was achieved by SAE-1DCNN (R^2 = 0.78, RMSE = 3.94%, RPD = 4.88) followed by 1DCNN (R^2= 0.73, RMSE = 5.43%, RPD = 3.67) proving the superiority of 1DCNN over FCNN in this study. These results supported the applicability of combined deep features extraction and regression methods for predicting SOC using high dimensional large-scale SSLs.

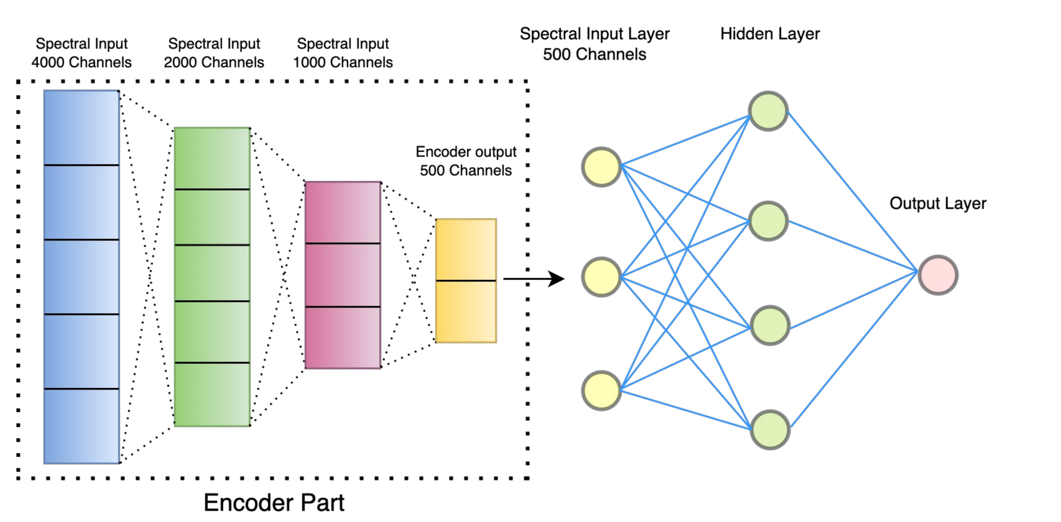
Method



Stacked autoencoder architecture

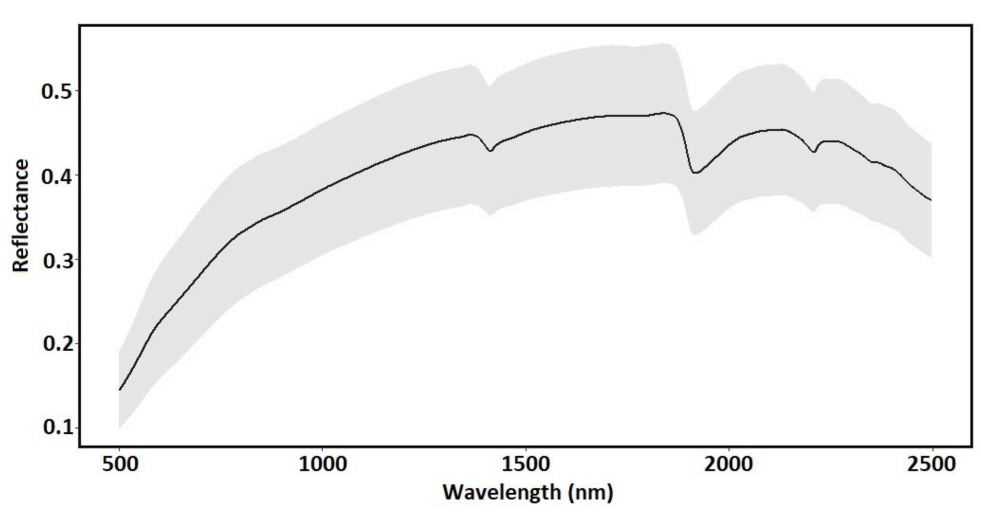


Stacked autoencoder-1D Convectional Neural Network

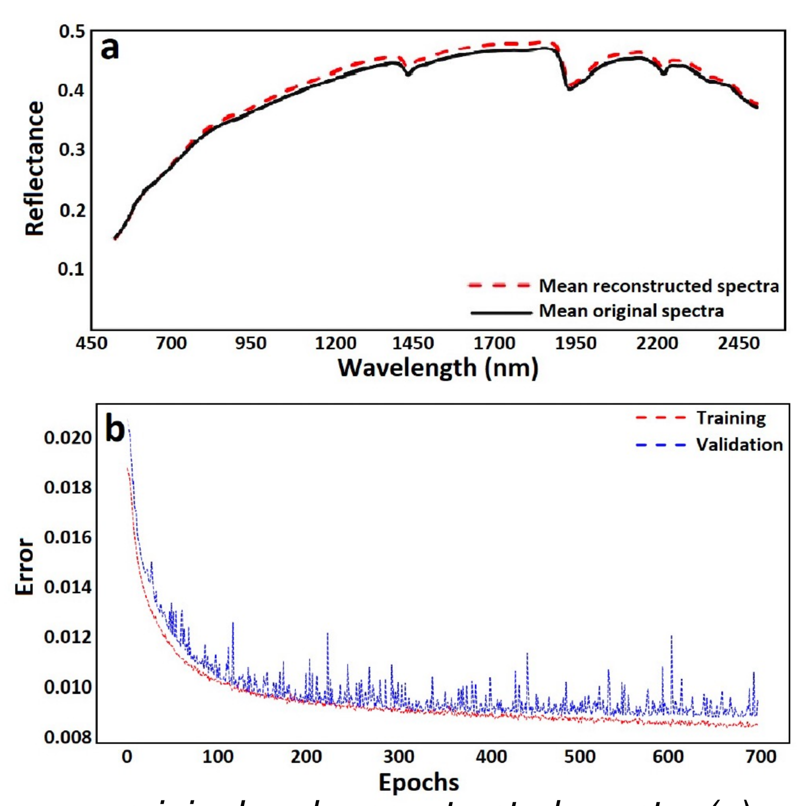


Stacked autoencoder fully connected neural Network

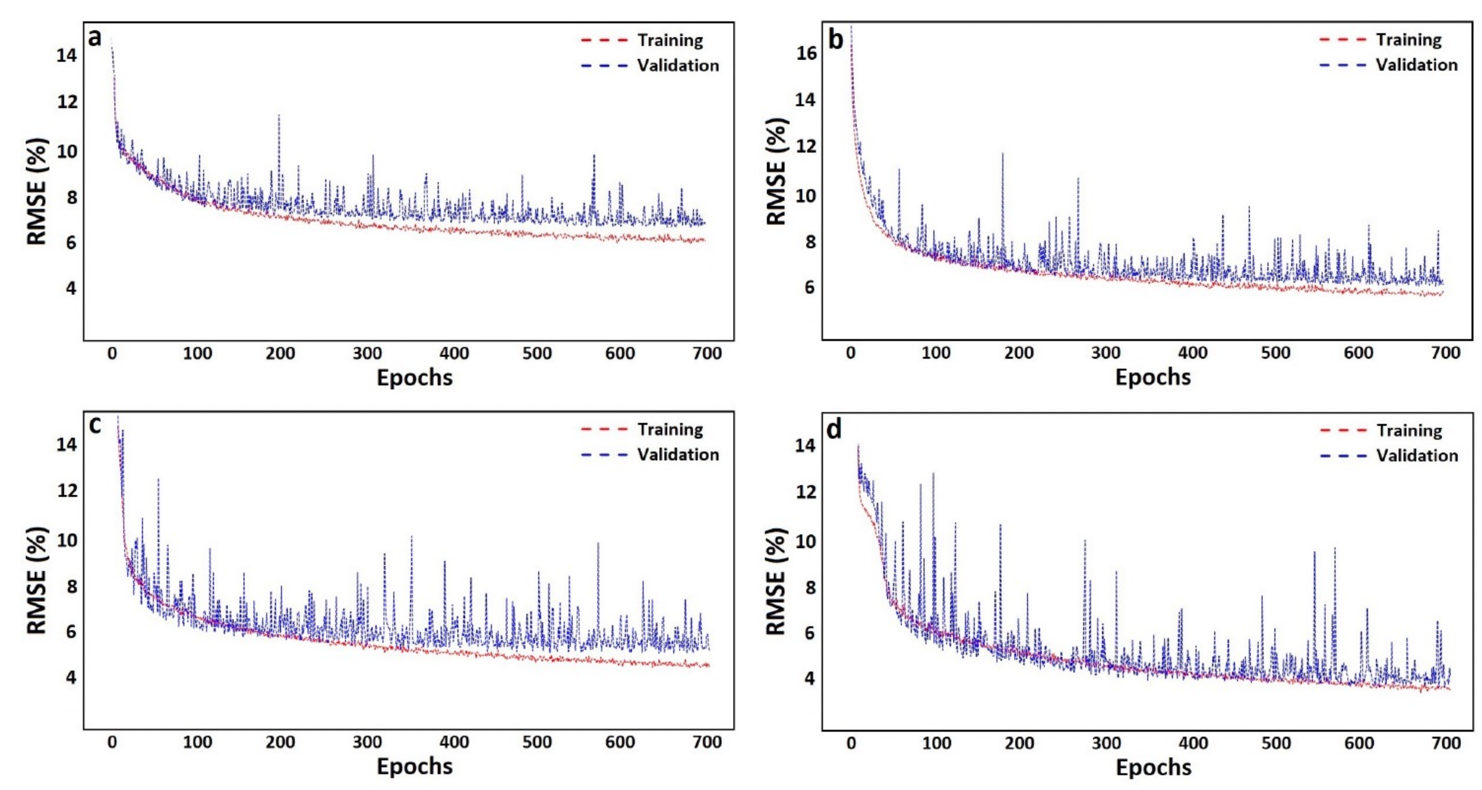
Results



The average reflectance of the whole dataset with variance



The mean original and reconstructed spectra (a), and Pre-training error (b)



Training/validation plots for FCNN (a), SAE-FCNN (b), 1DCNN (c), and SAE-1DCNN (d)

SOC prediction models performance based on different architectures

	EP	LR	BS	AF	R^2	RMSE	RPD
FCNN	700	-3	20	Leaky	0.625	6.715	2.966
SAE - FCNN	700	-3	20	Leaky	0.641	6.122	3.254
1DCNN	700	-3	16	Leaky	0.733	5.429	3.669
SAE - 1DCNN	700	-3	16	Leaky	0.784	3.935	4.884
RF	-	-	-	-	0.607	8.868	2.167

EP:Epochs, LR:learning rate, BS:Batch size, AF:Activation function

ACKNOWLEDGEMENT:

This work was supported by the European Union's Horizon H2020 research and innovation European Joint Programme Co-fund on Agricultural Soil Management (EJP-SOIL grant number 862695) and was carried out in the framework of the STEROPES of EJP-SOIL.

EJP SOIL STEROPES