Sample-efficient modeling of physical soil properties using convolutional neural networks and transfer learning

Will Heitman[1], Abhinav Dhakal[2], Callie Simon[2], Gary Feng[1], Jingdao Chen[2]

1: USDA-ARS, Genetics and Sustainable Agricultural Research Unit, 810 Hwy 12 East, Mississippi State, MS 39762
2: Bagley College of Engineering, Mississippi State University, Mississippi State, MS 39762.

USDA Agricultural Research Service

BAGLEY COLLEGE OF ENGINEERING

INTRODUCTION

- Soil spectroscopy measures how much light is reflected by the soil when light is applied to it.
- Soil properties such as minerals, organic compounds, and water can be estimated using soil spectroscopy, making it more cost- and time-efficient than traditional lab techniques.
- Soil spectroscopy does well when predicting most chemical properties but works less well for modeling physical properties.
- In this work, we apply the latest advances in machine learning, including transfer learning and contrastive loss, to build a model that predicts the wilting point using a plain Vis-NIR spectroscopy scan.

OBJECTIVES

- Use a convolutional neural network to predict the permanent wilting point from spectral data.
- Compare the performance of neural network regression methods against baseline random forest and Cubist methods.
- Use semi-supervised learning to overcome the challenge of unlabelled data.

MATERIALS AND METHODS

- Using Vis-NIR spectroscopy, three scans of each soil sample were taken, with wavelengths ranging from 350 nm to 2500 nm.
- Total of 225 labelled soil samples were taken from different locations in the state of Mississippi.
- We studied two kinds of neural networks: the Multi-Layer Perceptron (MLP) and the Convolutional Neural Network (CNN).
- An ablation study was conducted with the following components:

1. PCA: Principal Component Analysis, reducing 2150 features to just 80 important features.
2. Contrastive: Semi-supervised learning for using unlabelled data.
3. Backbone: Pre-trained the model on the OSSL dataset to predict % clay and bulk density.
4. Augmentation: Created different variations of the spectra data using Perlin and Gaussian noise.

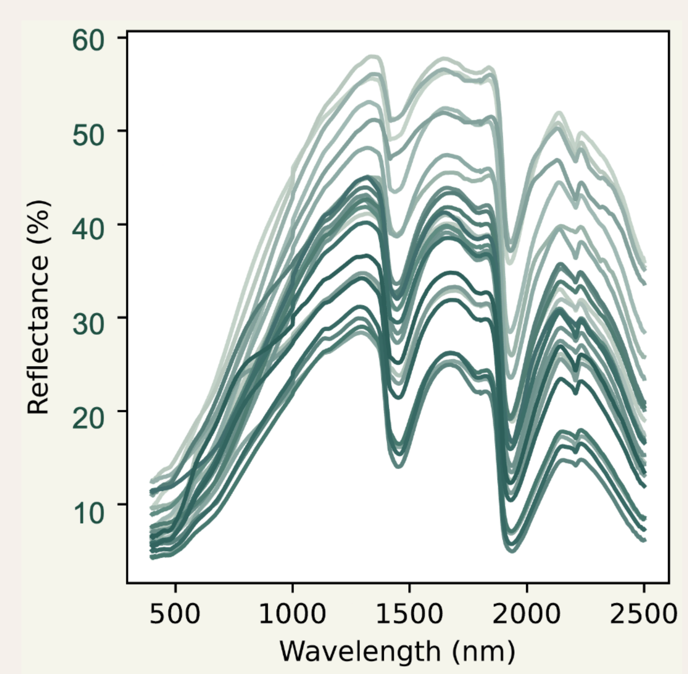

Figure 1: *The 99 train spectra, colored by wilting point from low to high.*

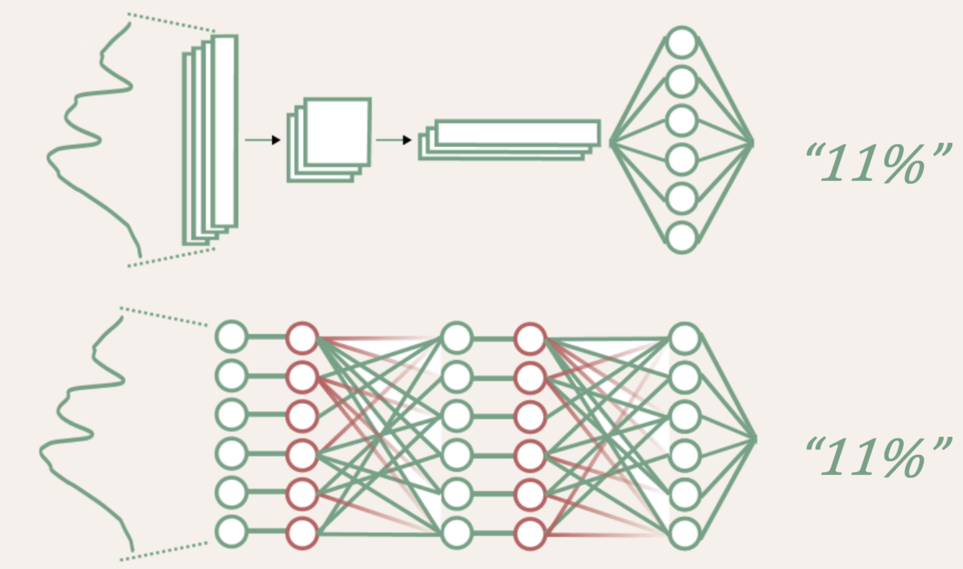

Figure 2: *Our CNN (top) and MLP with dropout (bottom) converts a raw spectral scan into a wilting point estimate.*

RESULTS AND CONCLUSION

Architecture	PCA	Contrastive	Backbone	Augmentation	Best RMSE (%water)	Final RMSE (%water)	Test R2
Cubist		-	✓	-	-	3.883	-0.148
Cubist	✓	-	✓	-	-	3.740	-0.373
Random Forest		-	✓	-	-	3.582	-0.066
Random Forest	✓	-	✓	-	-	3.221	-0.058
CNN		✓	✓	✓	2.394	2.416	0.200
CNN			✓	✓	2.305	2.309	0.268
MLP	✓	✓		✓	2.296	2.325	0.259
MLP	✓		✓	✓	2.289	2.402	0.209
MLP	✓			✓	2.287	2.408	0.205
MLP			✓		2.277	2.426	0.193
MLP		✓		✓	2.264	2.293	0.279
MLP				✓	2.261	2.315	0.265
MLP			✓	✓	2.244	2.264	0.297
CNN					2.238	2.258	0.301
CNN			✓		**2.237**	**2.258**	**0.301**

- ❖ Our model cut the error of the Cubist model by 40%, from 3.74% to 2.26%.
- ❖ Maximizing sample efficiency enables soil scientists to develop new models quickly and economically.
- ❖ Spectroscopy can replace wet lab tests so that farmers get physical properties readings in seconds instead of days.

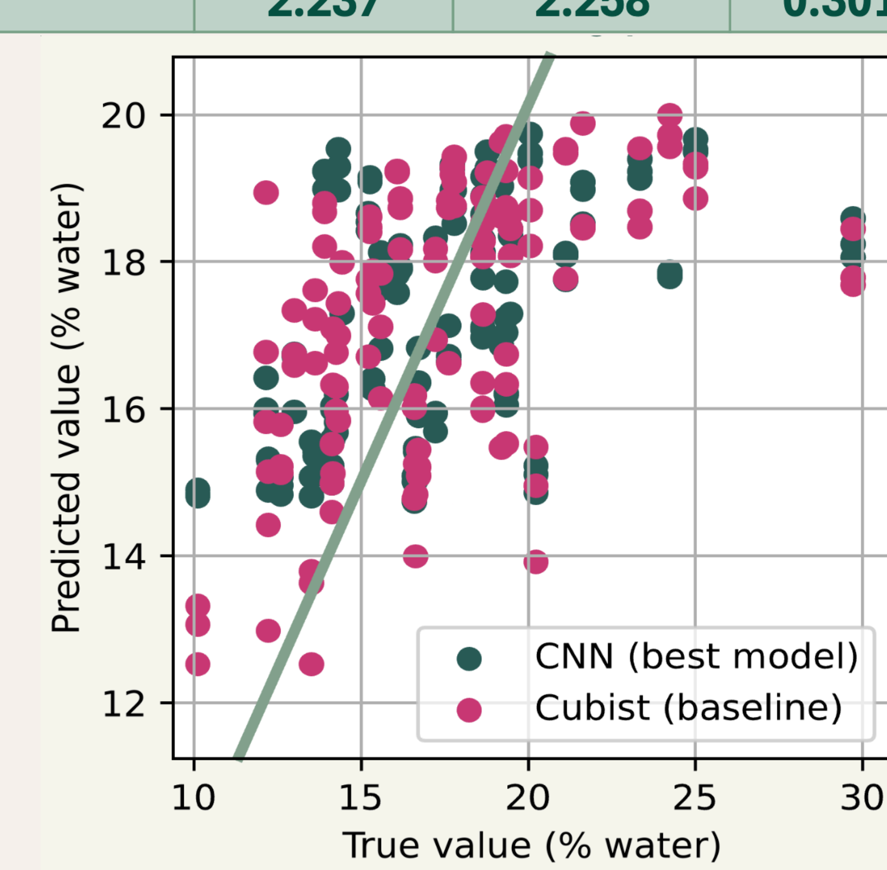

Figure 3: Demonstration of *soil spectroscopy using Fieldspec spectroradiometers at USDA-ARS, Starkville.*

FUTURE WORKS

- ❖ This research is part of a larger project with the long-term objective of using robotics to perform in-field scanning and analysis of soil properties.
- ❖ The findings from this research will be used to further experiment with other semi-supervised learning methods, such as contrastive learning and transfer learning.
- ❖ The findings will also be used to further research on predictive modelling of other soil properties, such as available water content, bulk density, total soil porosity, etc.

REFERENCES

- ❖ Chen, Ting & Kornblith, Simon & Swersky, Kevin & Norouzi, Mohammad & Hinton, Geoffrey. (2020). Big Self-Supervised Models are Strong Semi-Supervised Learners.
- ❖ Kirkham, M. B. (2014). *Principles of soil and plant water relations*. Elsevier.
- ❖ Wang, X., Zhang, M.-W., Guo, Q., Yang, H.-L., Wang, H.-L., & Sun, X.-L. (2023). Estimation of soil organic matter by in situ Vis-NIR spectroscopy using an automatically optimized hybrid model of convolutional neural network and long short-term memory network. *Computers and Electronics in Agriculture*, *214*, 108350–108350.

ACKNOWLEDGEMENT

- ❖ This research was supported by the Mississippi Soybean Promotion Board (62-2024)
- ❖ This research was supported by USDA-ARS AI-COE/SCINet internship.

MSPB