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

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

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

Introduction

- **Spectroscopy** can model soil properties by shining light on a sample and measuring the light that reflects (how "shiny" the soil is).
- Spectroscopy works well for measuring chemical properties, like the amount of carbon or the presence of chemical pollutants.
- …But it works less well for modeling physical properties, where the link between light reflectance and physical properties is relatively weak.
- In this work, we apply the latest advances in machine learning, including **transfer learning** and **contrastive loss**, to build a model that predicts wilting point using a plain Vis-NIR spectroscopy scan.
- We compare our model with Random Forest regressors and Cubist models (the SOTA).

Challenges

- This work addresses two challenges at once.
- First, there is a very weak correlation between reflectance and water retention. Any model we build **must be sensitive enough to find the nuances** (nonlinearities) in the spectroscopy scans.
- Second, we only had access to 225 labeled soil samples. Most studies train on tens of thousands, if not hundreds of thousands, of samples! Our model therefore needs to learn the most that it can from each sample we give it.

Ablation

- Studied two kinds of neural networks: the Multi-Layer Perceptron (MLP) and the Convolutional Neural Network (CNN).
- Pre-trained a backbone on the OSSL dataset to predict % clay and bulk density.
- Augmented spectra using Perlin and Gaussian noise.
- Added consistency regularization.

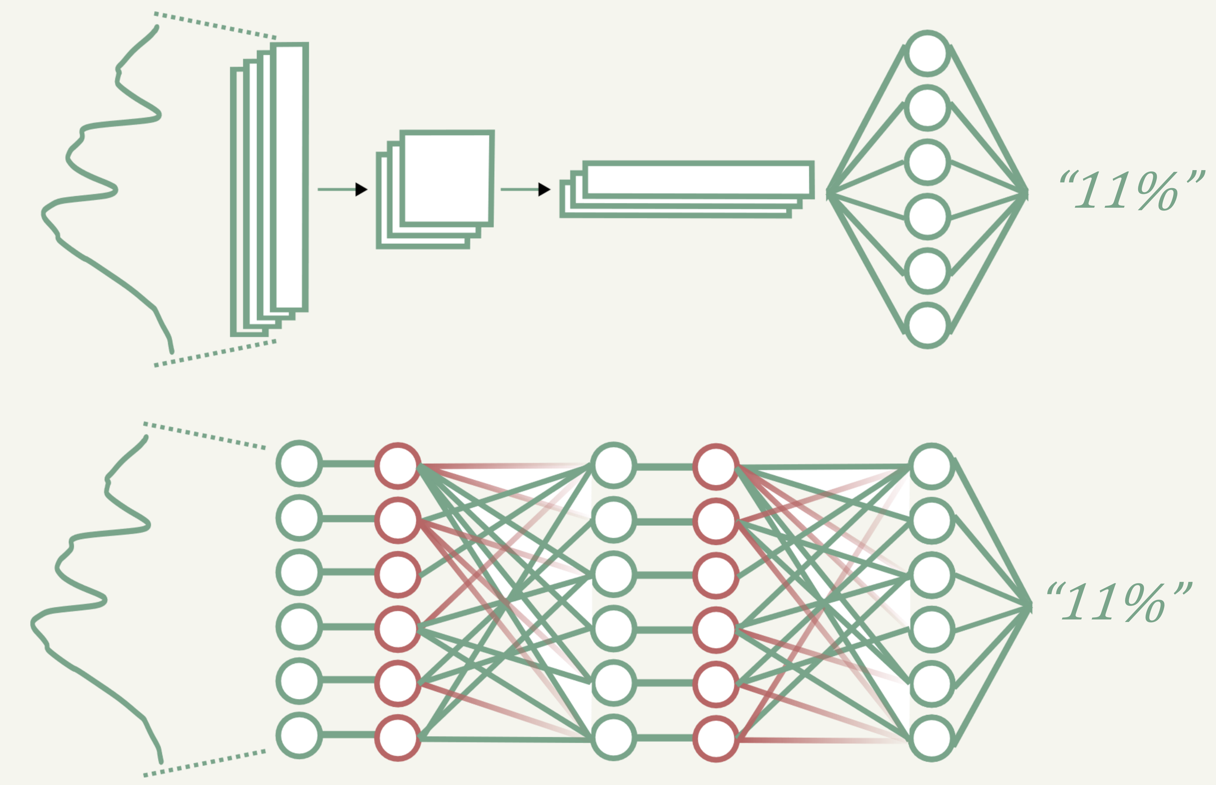

Figure 1: *Our CNN (top) MLP with dropout (bottom) convert a raw spectral scan into a wilting point estimate.*

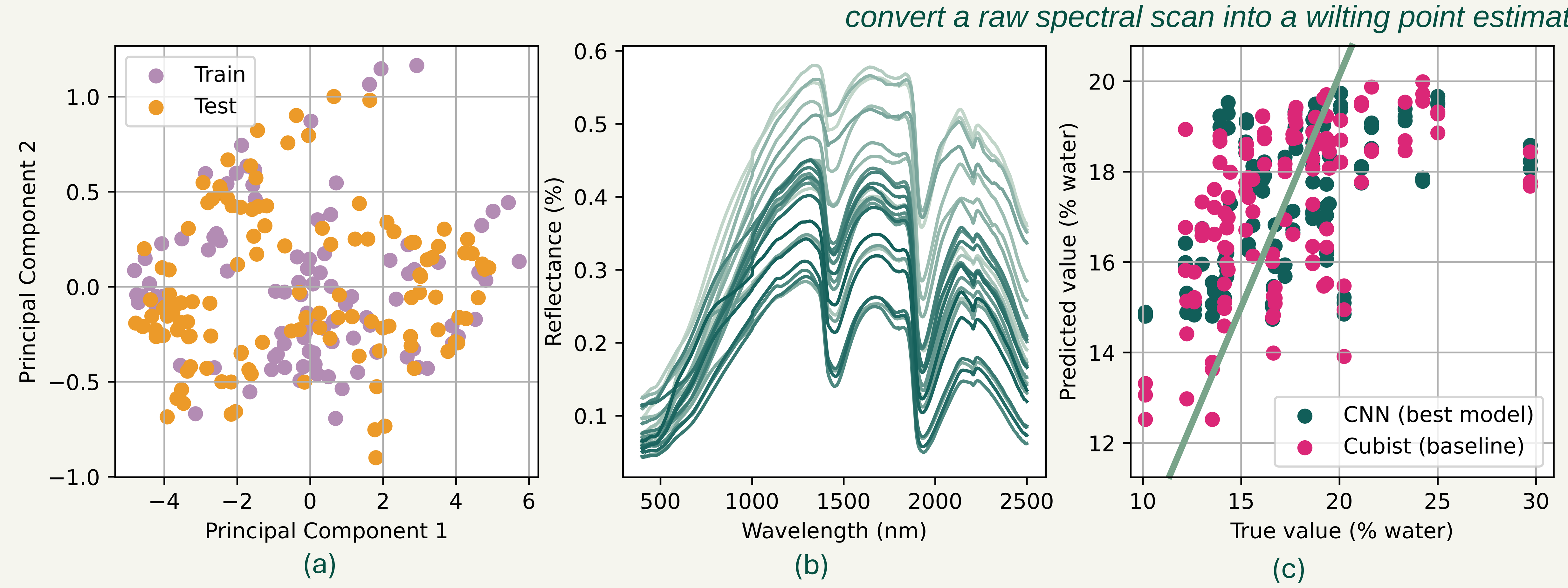

Figure 2: *(a) Distribution of the train and test data in the first two PCs. (b) The 99 train spectra, colored by wilting point from low to high. (c) True vs predicted wilting points in our best model and the baseline.*

Results

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**

Conclusion

- Our model achieved superior results than the previous SOTA on limited datasets.
- Maximizing sample efficiency enables soil scientists to develop new models quickly and economically.

Figure 3: *A scientist at the Woodwell Climate Research Center collects a spectroscopy scan in the field.*

Future work

- Leverage unlabeled spectra using semi-supervised learning techniques, such as temporal ensembling.
- Predict other kinds of soil physical indicators, such as available water content or field capacity.
- Deploy our model onto a robotic platform to perform automated field scanning.

References

[1] D. Wang et al., "A lightweight convolutional neural network for nicotine prediction in tobacco by near-infrared spectroscopy," Frontiers in Plant Science, vol. 14, May 2023. doi:10.3389/fpls.2023.1138693
[2]J. L. Safanelli, T. Hengl, J. Sandermanet L. Parente, "Open Soil Spectral Library (training data and calibration models)". Zenodo, Dec. 26, 2021. doi: 10.5281/zenodo.7599269.
[3] X. Wang *et al.*, "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*, vol. 214, p. 108350, Nov. 2023. doi:10.1016/j.compag.2023.108350

View code and further details: heit.mn/mas (or scan the QR code)