

Deep learning applications for evaluating taxonomic and morphological diversity in ferns

Alex White

Mike Trizna, Sylvia Orli, Eric Schuettpelz,
Paul Frandsen, Rebecca Dikow, Larry Dorr

Smithsonian Institution

National Museum of Natural History – Dept. of Botany

Data Science Lab



Where are we at?

U.S. National Herb: 2 million specimens are digitized

Fern collection is completely digitized:

215,000 specimens with global representation

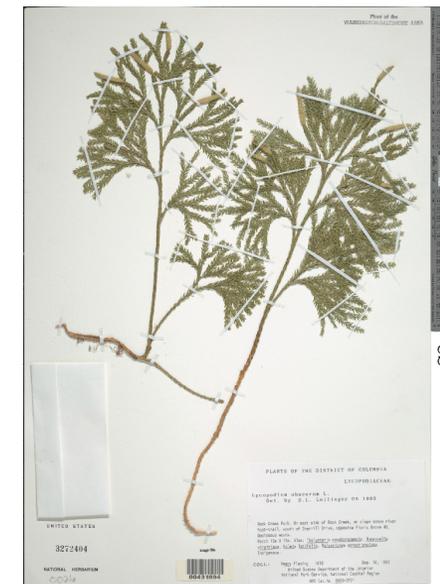
86 genera with >500 specimens

Past work focused on binary classification:

Mercury staining and 2 genera of clubmoss

Build a more robust taxonomic classification model?

Will taxonomic classification inform us about ecological differences in shape?



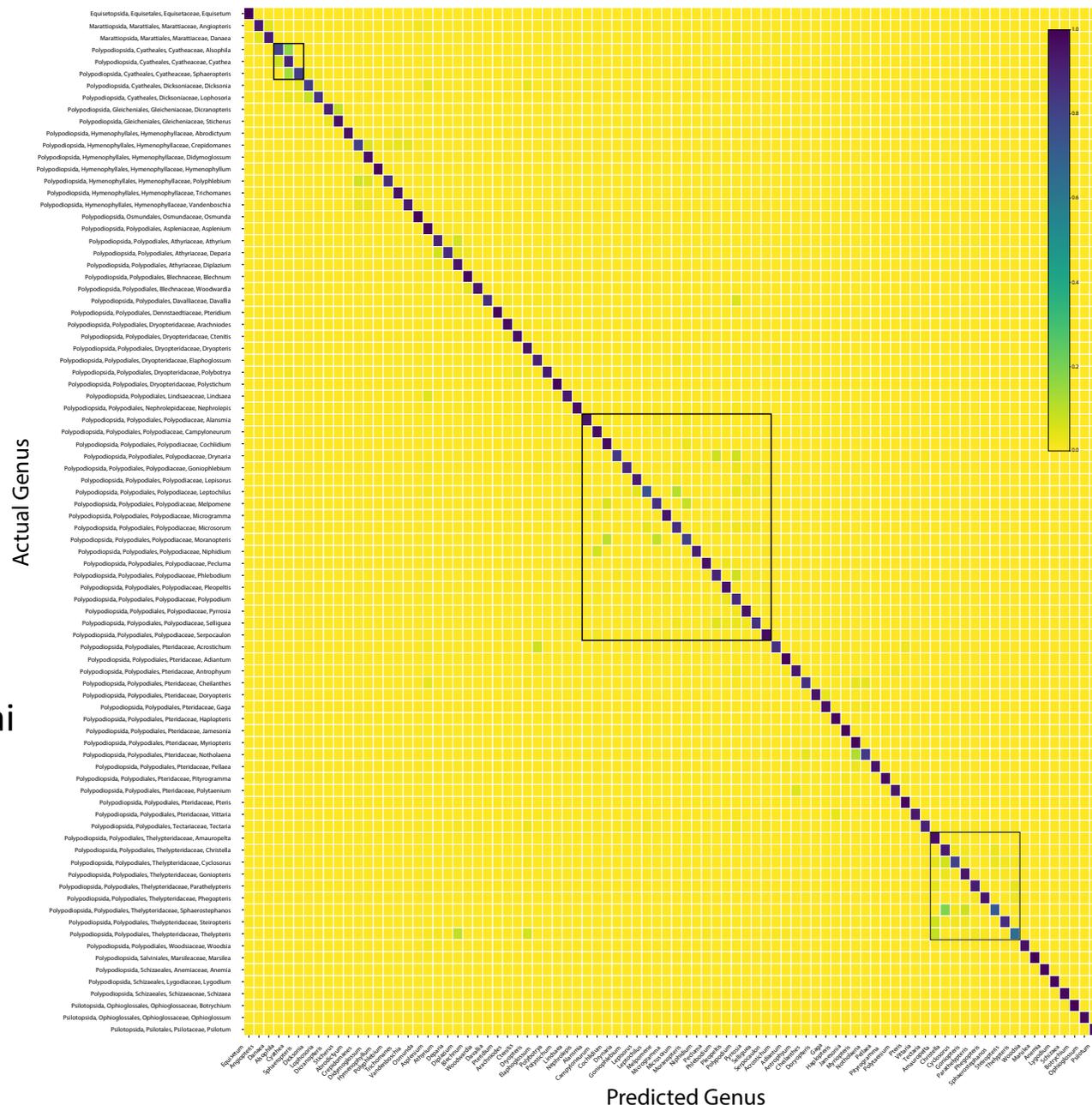
Where are we at?

175,000 specimens in 86 genera (>500)
Some genera >15,000 specimens

97% accuracy for 86 genera

The taxonomic classifier:

- CNN with res-net 101 backbone
- Built in PyTorch
- 256 x 256 images
- Trained using the 1cycle policy (Smith, 2018) in fastai
- Transforms = flipping, rotating, zooming, lighting



Where are we going?

Given the model accuracy, can the features of the model tell us something about morphological similarity?

Take the embedding vectors from our classifier and use UMAP for dimensionality reduction

Can this 2D space be thought of as a morpho-space?

Problems: coloration, other materials on the sheet

