# Phenotyping Alfalfa (*Medicago sativa* L.) Root Structure Architecture Via Integrating Confident Machine Learning with ResNet-18

*Brandon Weihs, USDA-ARS*
*Zhou Tang, Washington State University*
*Zezhong Tian, University of Wisconsin - Madison*
*Jo Heuschele, USDA-ARS*
*Aftab Siddique, Fort Valley State University*
*Thomas Terrill, Fort Valley State University*
*Zhou Zhang, University of Wisconsin - Madison*
*Larry York, Oak Ridge National Laboratory*
*Zhiwu Zhang, Washington State University*
*Zhanyou Xu, USDA-ARS*

Root system architecture (RSA) is of growing interest in implementing plant improvements with belowground root traits. Modern computing technology applied to images offers new pathways forward to plant trait improvements and selection through RSA analysis. However, image label noise reduces the accuracies of models. This study utilized an AI model (ResNet-18) capable of classifying RSA directly from alfalfa root crown images. Images were compared with different model outputs (Random Forest and ResNet-18) with manual root classifications, then tested confident machine learning (CL) and reactive machine learning (RL) methods to minimize the effects of subjective labeling and improve labeling and prediction accuracies. The CL algorithm modestly improved the Random Forest model's overall prediction accuracy of the Minnesota data set (1%) while larger gains in accuracy were observed with the ResNet-18 model results. The ResNet-18 cross-population prediction accuracy was improved (~8 to 13%) with CL compared to the original/uncorrected datasets. Training and testing data combinations with the highest accuracies (86%) resulted from the CL and/or RL corrected datasets for predicting taproot RSAs. Similarly, the highest accuracies achieved for the intermediate RSA class resulted from corrected data combinations. The highest overall accuracy (~75%) using the ResNet-18 model involved CL on a pooled dataset containing images from both sample locations. This study concluded that the ResNet-18 DNN prediction accuracies of alfalfa RSA images are increased when CL and RL are employed. By increasing the dataset to reduce overparameterization while concurrently finding and correcting image label errors, it is demonstrated here that accuracy increases as much as ~11-13% can be achieved with semi-automated, computer-assisted preprocessing and data cleaning (CL/RL). This research, while still fledgling, moves alfalfa breeding/selection closer to objective RSA phenotyping which aims to remove human error and bias by incorporating machine learning into plant selection and breeding tasks as well as increase the speed of AI-driven RSA analyses via directly inputting images (pixel data) into models such as ResNet-18 instead of the typical feature-data-as-input process.