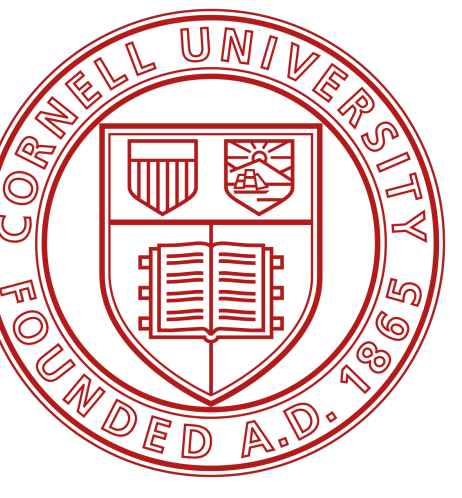




Prospector: An Android application for NIRS-based phenotyping of cassava quality traits



Jenna Hershberger¹, Ugochukwu Ikeogu¹, Williams Esuma², Robert Kawuki², Ismail Rabbi³, and Michael A. Gore¹

¹Plant Breeding, Cornell University, USA; ²NaCRRI, Uganda; ³IITA, Ibadan, Nigeria;

Background

- High root dry matter content (DMC) is a primary breeding objective in most cassava breeding programs
 - Traditional measurement requires considerable time and/or effort¹
 - Near-infrared spectroscopy (NIRS) has been shown to be highly predictive of dry matter in cassava roots²
 - Tested spectrometers are prohibitively expensive
- SCiO is an affordable, handheld NIR spectrometer that connects to a customizable smartphone app for data collection



SCiO specs

- Portable, handheld
- \$300/device
- Range: 740-1070nm
- Bluetooth connection to smartphone application



Objectives

- Evaluate an affordable spectrometer for ease of use and predictive ability in cassava DMC
- Develop robust and accurate prediction models for DMC in cassava using SCiO spectral data
- Integrate prediction models into the PhenoApp suite for distribution to cassava breeders

Current status

1. Calibration set development

- May-June 2018 harvest data from IITA has been processed, including scans and oven DMC data from two breeding populations with two reps each (total = 203 plots, 1434 roots)
- NaCRRI harvest data, including multiple East African breeding populations, is currently being processed (298+ plots)

2. Initial prediction model development

- Models developed for the 2018 IITA calibration set using every combination of 11 preprocessing methods and 3 model types
- R^2 was calculated as the squared correlation of predicted vs observed values using the test sets from these models (see right)

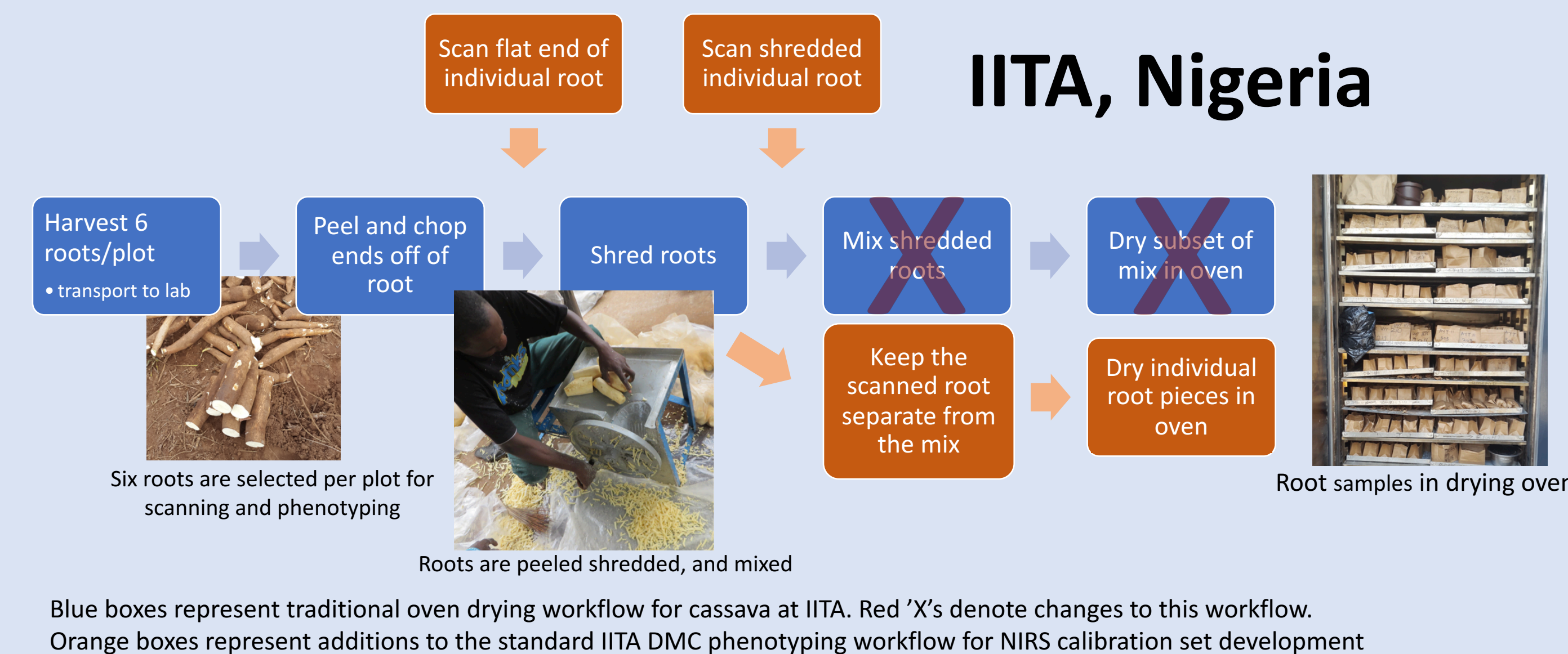
3. Model evaluation and optimization

- Optimal model type and parameters as determined for the IITA calibration set: PLSR model with Savitzky-Golay filter, second derivative, and window size of 5

4. App integration

- Prototype app is under development by the PhenoApps team

1. Calibration set development



NaCRRI, Uganda



2. Initial prediction model development

Preprocessing

- Smoothing (scatter reduction)
 - Standard normal variate (SNV)
 - First derivative (D1)
 - Second derivative (D2)
- Noise reduction
 - Savitzky-Golay (SG)
 - Window size 5 (W5)
 - Window size 11 (W11)

Model types

- Partial Least Squares Regression (PLSR)
- Random Forest (RF)
- Support Vector Machine (SVM)

Tuning and training

- 5-fold cross validation (hyperparameter tuning)
- 70% training / 30% test (model selection)

R^2 for preprocessing and models, 2018 IITA calibration set

Preprocessing technique	PLSR	RF	SVM
Raw data	0.78	0.57	0.63
SNV	0.75	0.65	0.4
SNV.1D	0.77	0.73	0.37
SNV.2D	0.78	0.7	0.67
D1	0.75	0.66	0.4
D2	0.77	0.65	0.68
SG	0.78	0.58	0.62
SNV.SG	0.71	0.6	0.54
SG.D1.W5	0.76	0.66	0.49
SG.D2.W5	0.81	0.69	0.68
SG.D1.W11	0.78	0.7	0.45
SG.D2.W11	0.76	0.64	0.34

The 2018 IITA calibration dataset of 203 plots (n= 1434 roots), each with root scans and oven DMC measurements, was split into 70% training and 30% test sets. The training set was used to tune PLSR, RF, and SVM hyperparameters under 5-fold cross validation. DMC predictions were generated for the test sets using these models, and correlations between the predictions and observed oven measurements were squared and averaged to produce a representative R^2 value for each model (above).

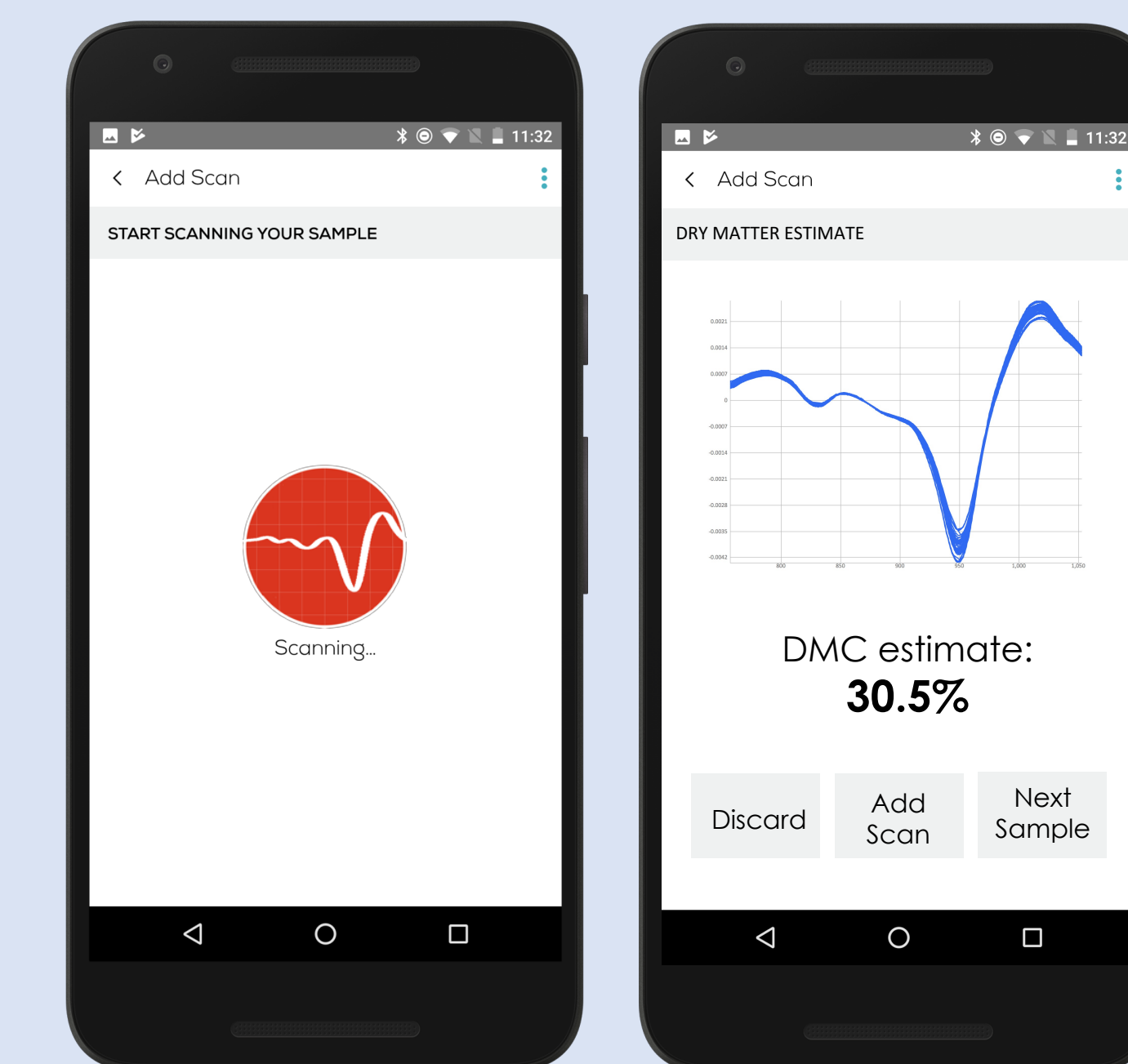
3. Model selection and optimization

Determine best model type and hyperparameters for each calibration set

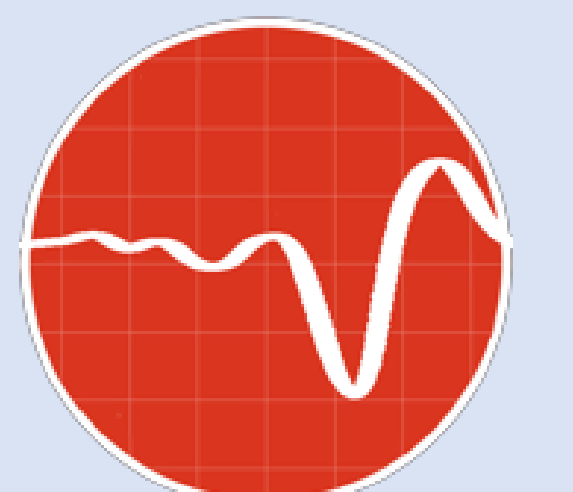
- Iterate through steps 1 and 2, adding calibration data that are representative of the samples to be phenotyped with the prediction models

4. App integration

- Prediction models will be distributed in Android app form to breeding programs as Prospector, part of the PhenoApps suite



PHENOAPPS



Prospector

Conclusions

- The range of wavelengths measured by SCiO are predictive of DMC in cassava roots
- Integration of a NIRS-based DMC phenotyping method into existing phenotyping workflows is both feasible and desired by the cassava breeding community

Future work

Prediction model optimization and evaluation for DMC with SCiO

- Continue large-scale calibration model establishment
- Incorporate NaCRRI Uganda data
- Determine effectiveness of models across environments and cassava breeding populations

SCiO app integration

- Improve user interface and incorporate prediction models
- Beta-test with collaborators at IITA and NaCRRI

References and acknowledgements

¹ Fukuda, W.M.G. et al. (2010). *Intl Inst Trop Agric*. 19.
² Sánchez, T. et al. (2014). *Food chemistry*. 151:444-451.

Cassava breeding teams at NaCRRI, Uganda and IITA, Nigeria
 PhenoApps project team at KSU and Makerere University

This material is based upon work supported by the National Science Foundation (NSF) under NSF-BREAD Grant No. 1543958. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

