



## Prediction of Maize Genotype from Hyperspectral Scans of Seeds Using Deep Learning

*Jorge Gutierrez<sup>1</sup>, Katie Murphy<sup>1</sup>, Sam Kenney<sup>1</sup>, Julian Cooper<sup>2</sup>, Cory D. Hirsch<sup>2</sup>, Malia A.*

*Gehan<sup>1</sup>, Noah Fahlgren<sup>1</sup>*

*<sup>1</sup>Donald Danforth Plant Science Center, St. Louis, MO, USA*

*<sup>2</sup>University of Minnesota, Saint Paul, MN, USA*

**ORCID: 0000-0002-8656-1567**

**Keywords:** Hyperspectral Imaging, Deep Learning, Maize Genotype Identification

Grain and seed properties can be evaluated using near-infrared spectroscopy and other methods for post-harvest quality assessment. Hyperspectral imaging combines spectroscopy with spatial information, which provides additional features that may improve predictive models of seed traits. To assess the ability of deep learning models to use hyperspectral data for predicting phenotypes, we first aimed to predict the genotype of maize seeds. Previous work achieved high identification accuracy between a small set of genotypes using either RGB images or hyperspectral data, and we hypothesized that high spectral resolution (350-1000nm) hyperspectral data would outperform simple RGB data in our study. Our dataset consisted of hyperspectral images of maize seeds from 47 inbred lines, including the 26 NAM lines, with 96 individual seeds per genotype. We evaluated the difference in genotype identification accuracy using three different representations of the individual seed data: 1) using the whole scan, containing the reflectance at 580 different wavelengths, 2) using a subset containing the reflectance at 3 different wavelengths corresponding to a pseudo-RGB image, and 3) a gray-scale image derived from the pseudo-RGB image. We fine-tuned VGG11, a popular convolutional neural network, using 85% of the individual seed data for each of the representations. We obtained around 90% genotype prediction accuracy on the unseen data for both the whole scan and the pseudo-RGB data, and 72% genotype prediction accuracy using the gray-scale data. The results indicate that the shape and color information contained in RGB images might be sufficient for the task of maize seed genotype identification.