CROP PHENOTYPING THROUGH SPECTRAL CLASSIFICATION

General Assembly Data Science Immersive

Marta Ghiglioni

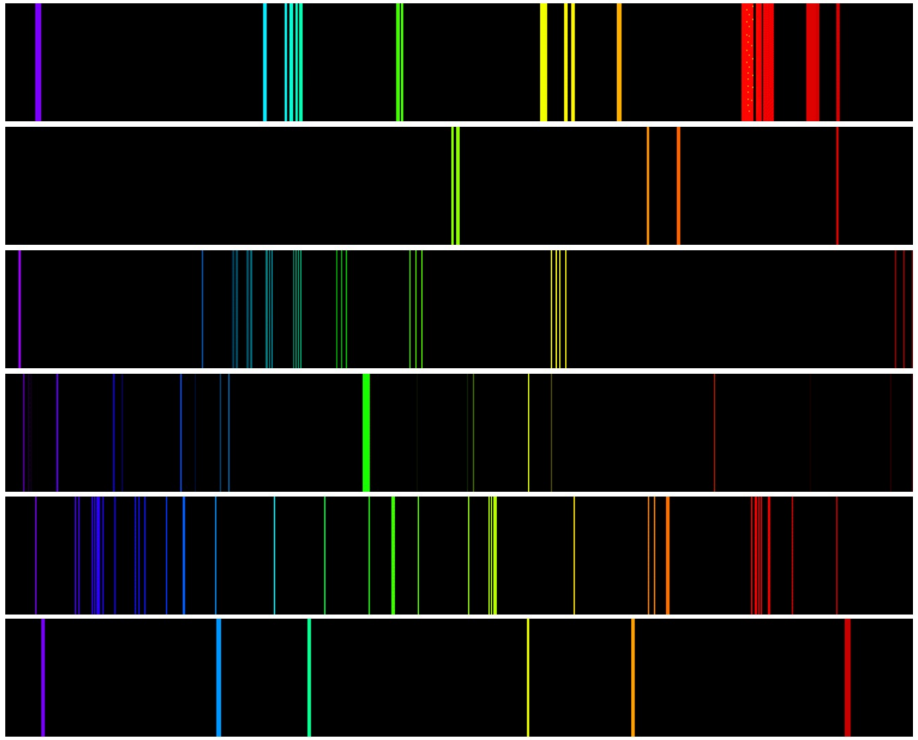
BACKGROUND

- Food security is one of the big challenges policy makers worldwide are facing.
- To feed 7.4+ billion people sustainably we need high quality data to inform policy makers to drive ultra precision farming at global scale.

HYPERSPECTRAL IMAGING

Solar Radiation Spectra

Spectral Signatures



(N) NITROGEN

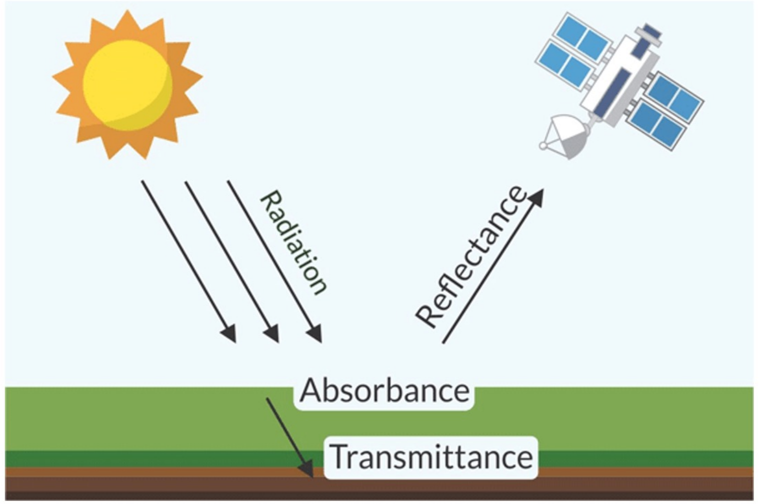
(P) PHOSPHORUS

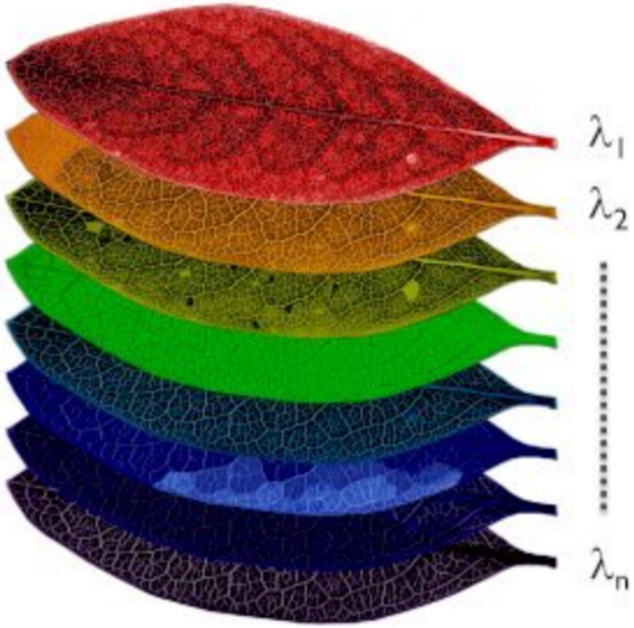
(K) POTASSIUM

(Mg) MAGNESIUM

(Ca) CALCIUM

(S) SULFURE

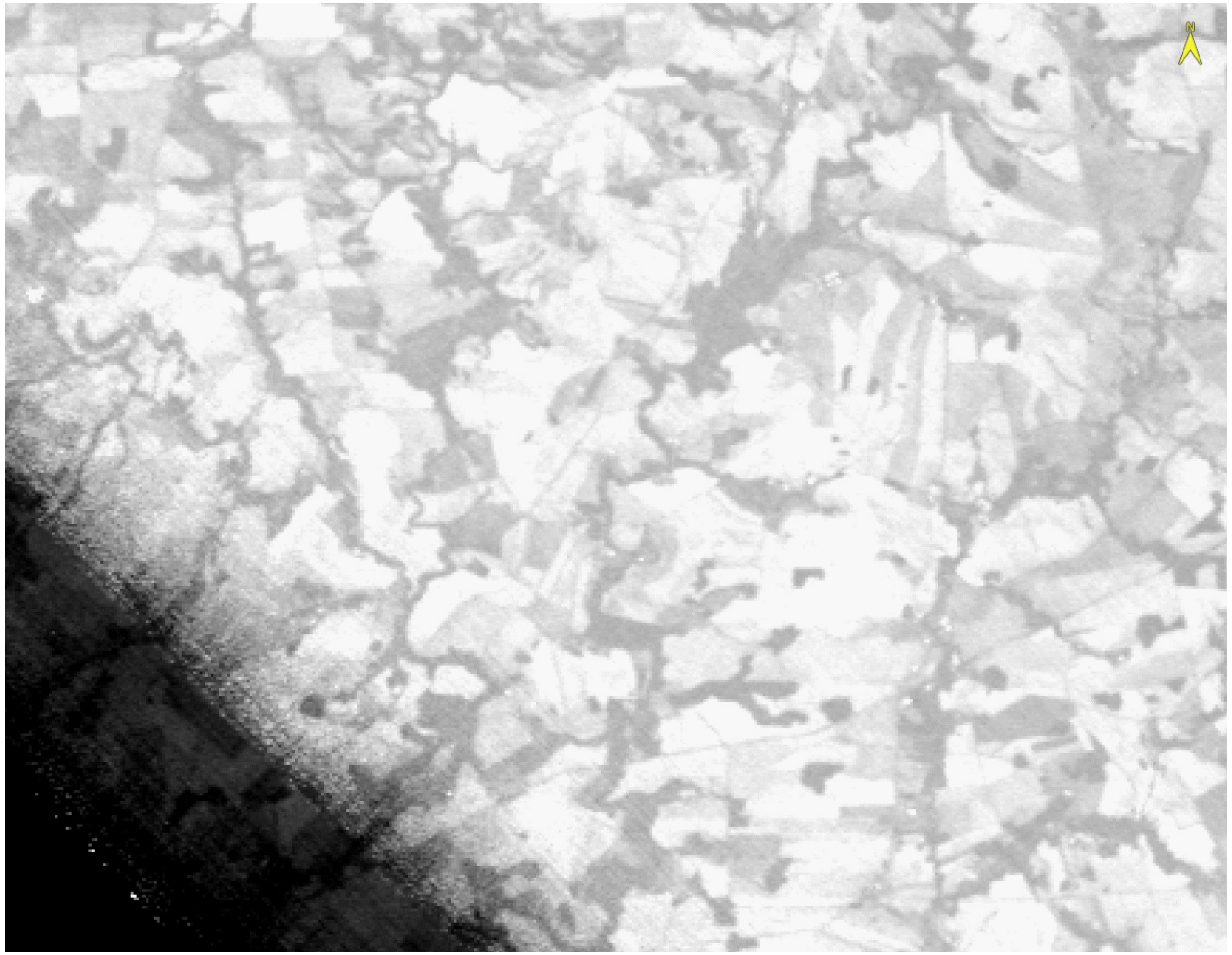




HYPERSPECTRAL IMAGING

Becoming a more popular data source because:

+ Scale access through low orbit satellite imaging
+ Investigating ways to automatize the insights extraction



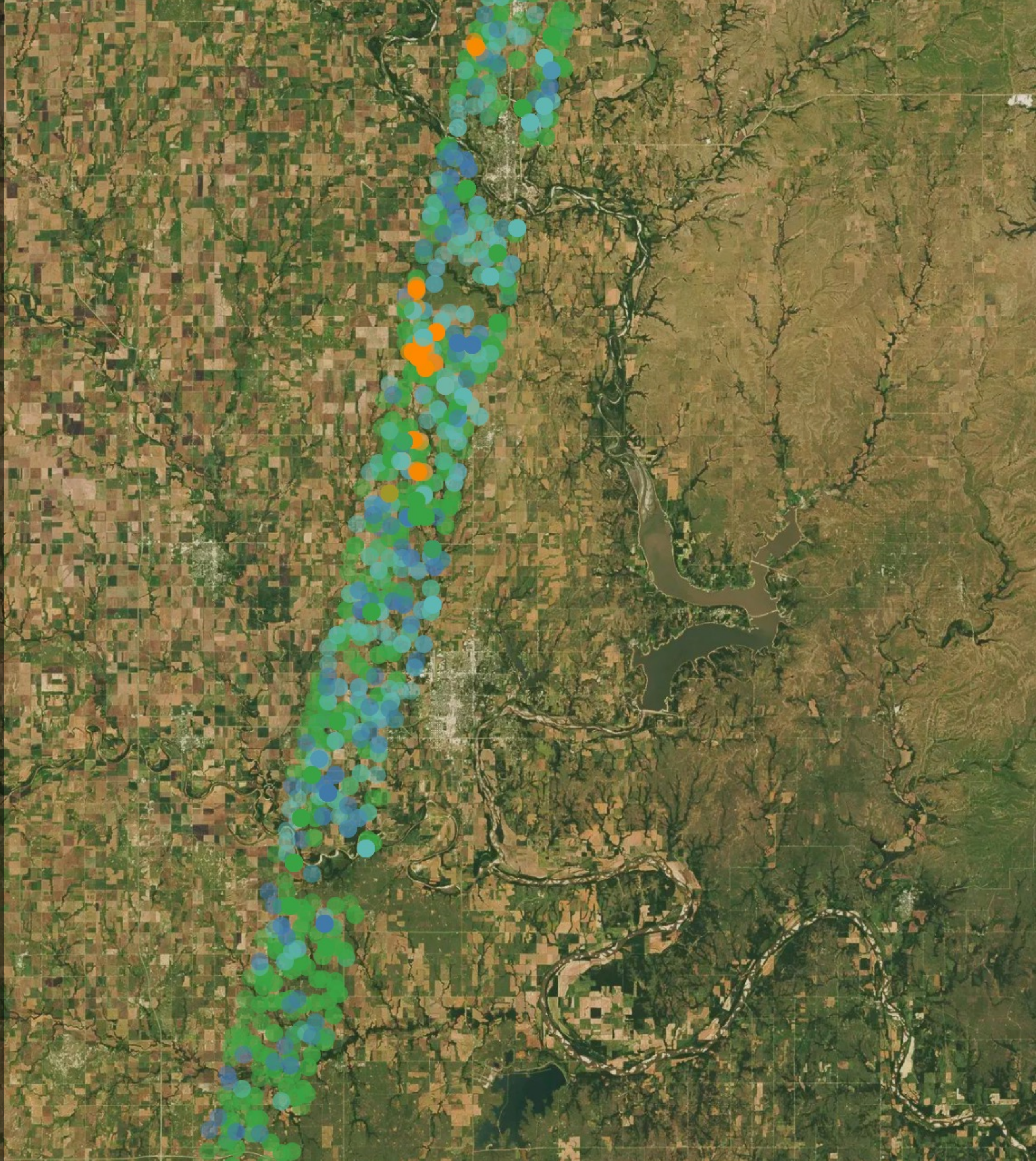
PROBLEM STATEMENT

+ Is it possible to identify crop types and their development stage through their spectral signature?

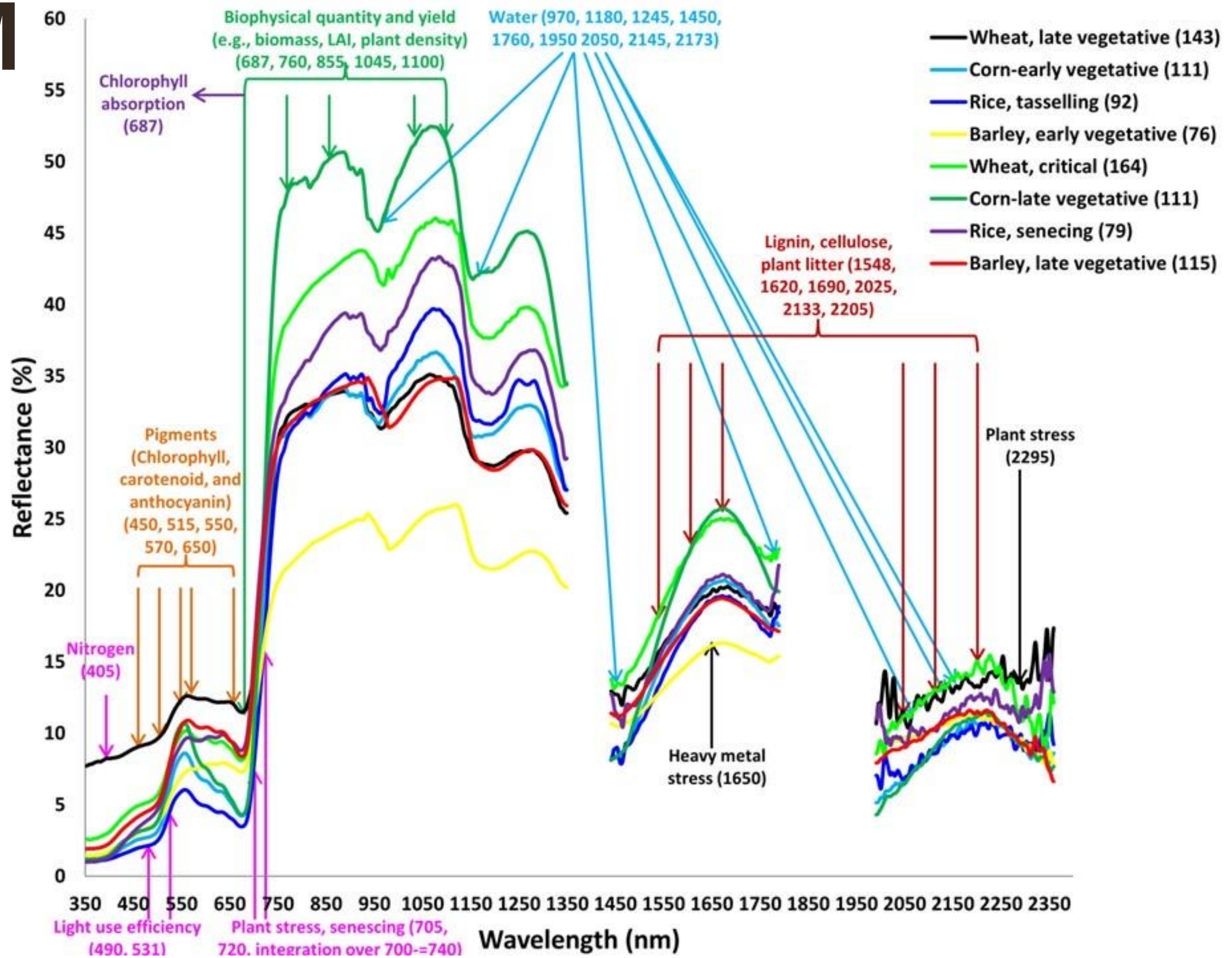
+ How can we use these information for food security?

GHISA

Global
Hyperspectral
Imaging
Spectroscopy of
Agricultural-Crops & Vegetation



CROP SPECTRUM



DATA SOURCES

FEATURES

- Hyperspectral images
  EO-1, Hyperion hyperspectral data
- 242 bands, each of 10 nm bandwidth in the 400-2500 nm range

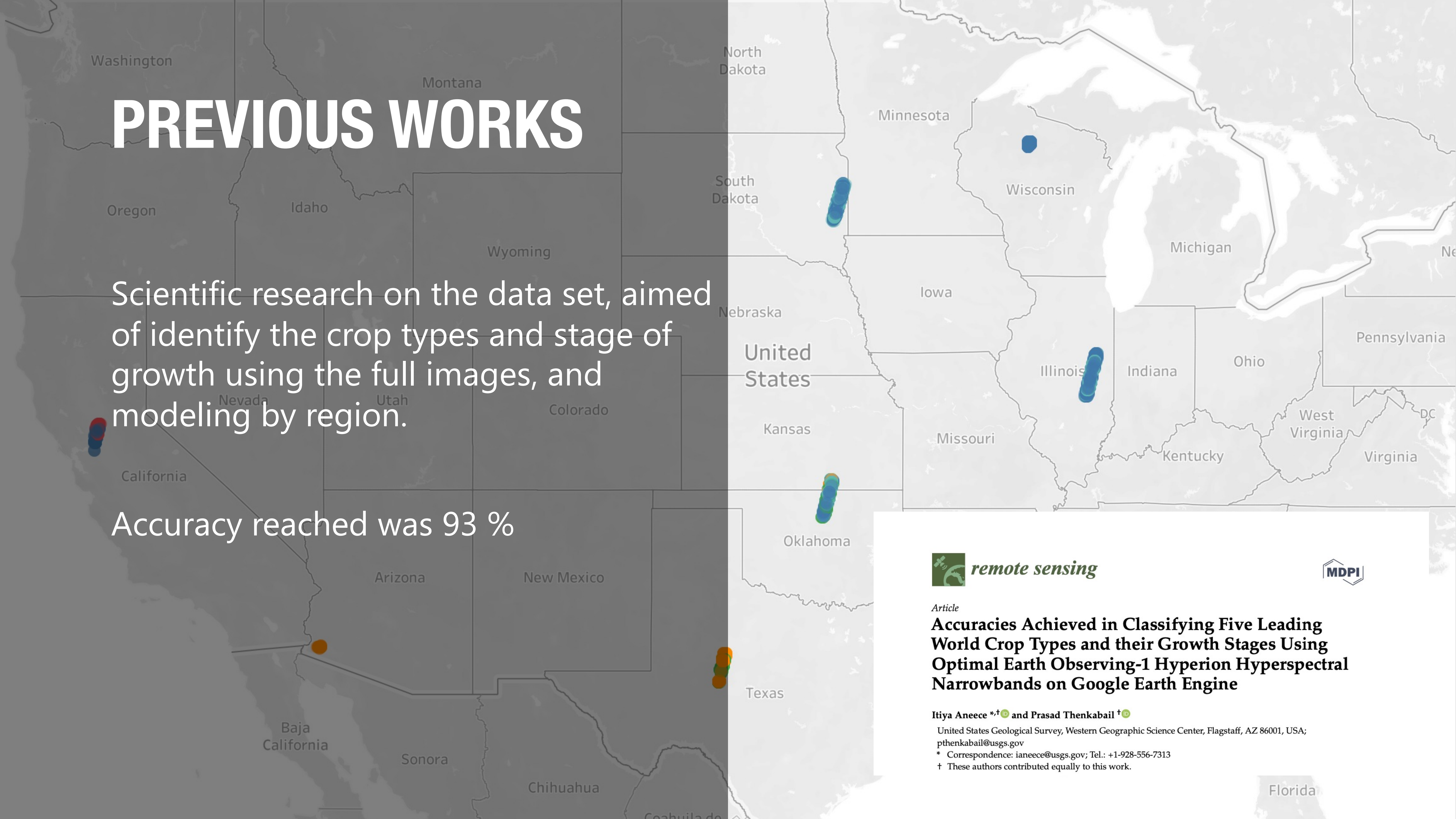
LABELS

- Crop type:
  USDA Cropland Data Layer
- Stage type:
  Center for Sustainability and the Global Environment (SAGE)

PREVIOUS WORKS

Scientific research on the data set, aimed of identify the crop types and stage of growth using the full images, and modeling by region.

Accuracy reached was 93 %



remote sensing

MDPI

Article

Accuracies Achieved in Classifying Five Leading World Crop Types and their Growth Stages Using Optimal Earth Observing-1 Hyperion Hyperspectral Narrowbands on Google Earth Engine

Itiya Aneece *,† and Prasad Thenkabail †

United States Geological Survey, Western Geographic Science Center, Flagstaff, AZ 86001, USA; pthenkabail@usgs.gov

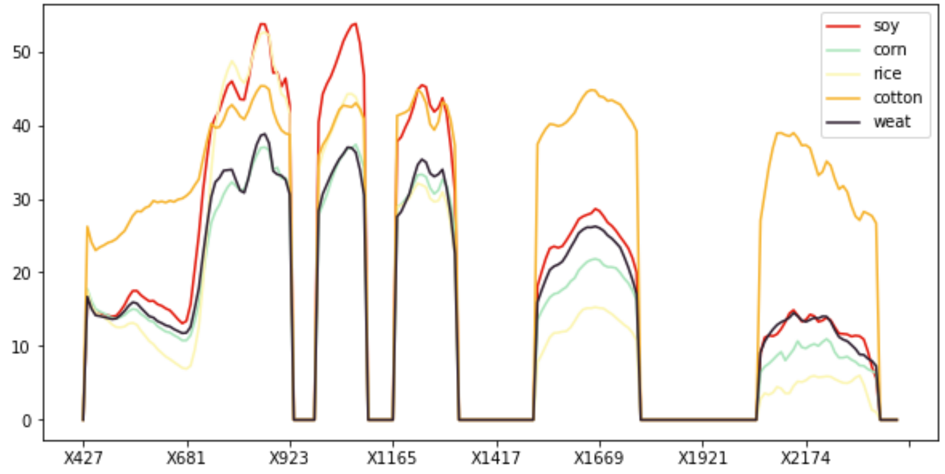
* Correspondence: ianeece@usgs.gov; Tel.: +1-928-556-7313

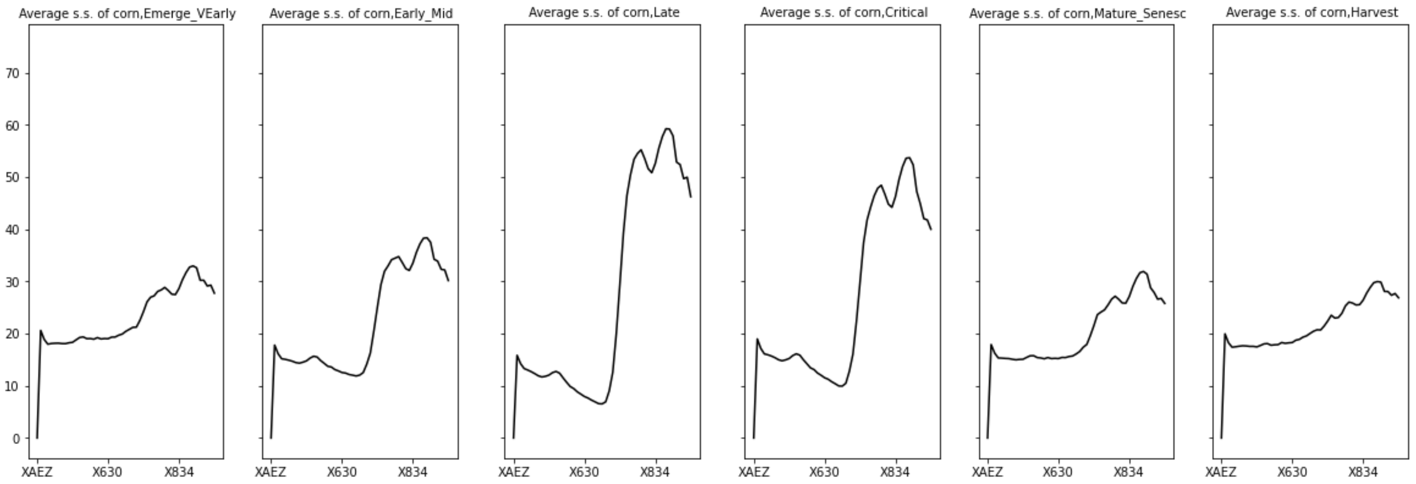
† These authors contributed equally to this work.

PRELIMINARY ANALYSIS

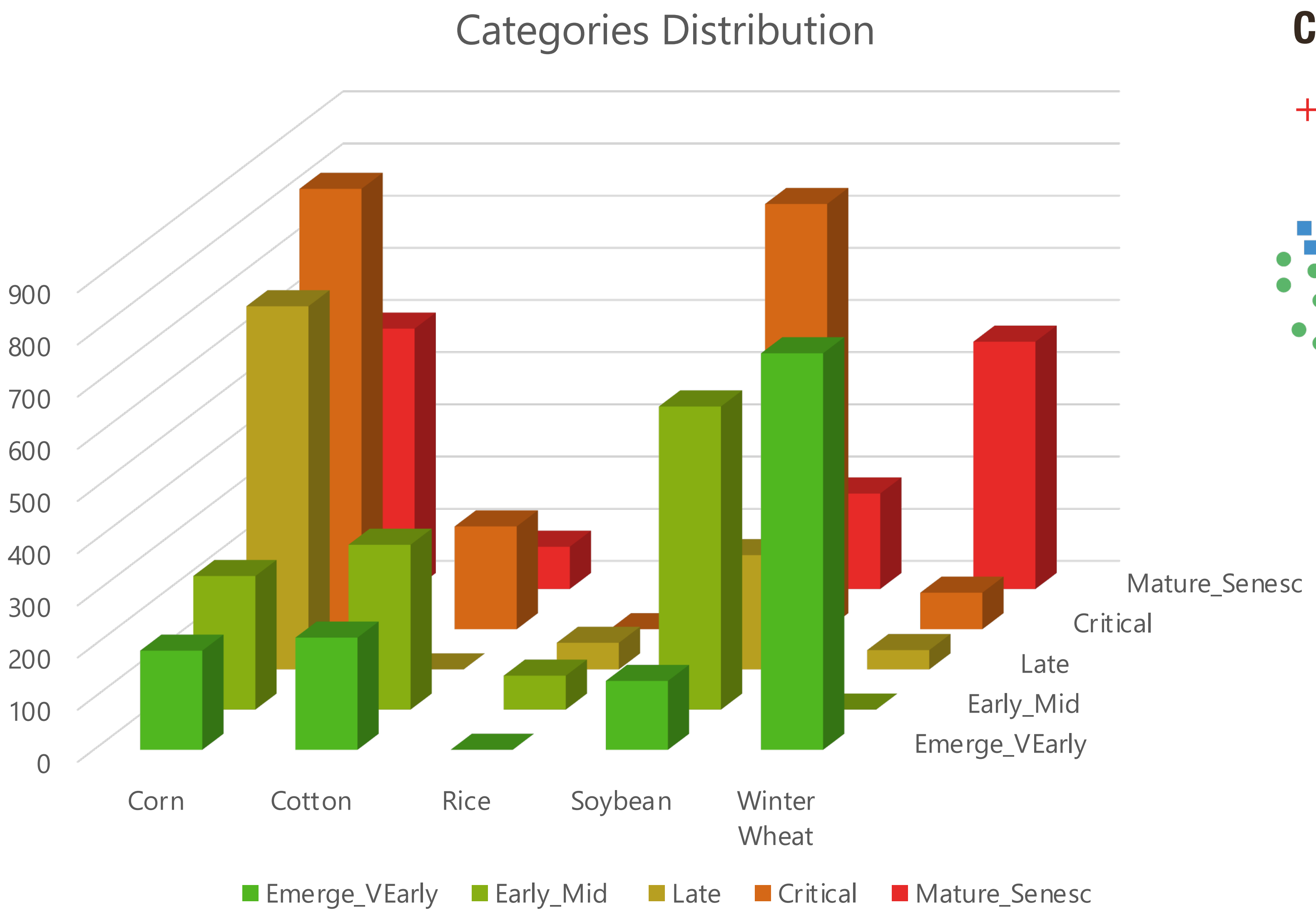
+ After reading research papers I decided to attempt to **classify crop type and stage using only their spectral signature**.

+ I also decided to focus my research on the **first portion of the spectrum** (bands 427 – 923). Indeed, these bands are more suitable for a larger scale of applications, easier to collect and have a higher resolution.



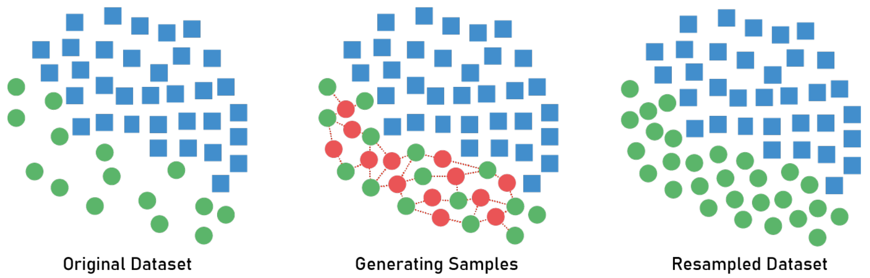


DATASET



CLASS UNBALANCE

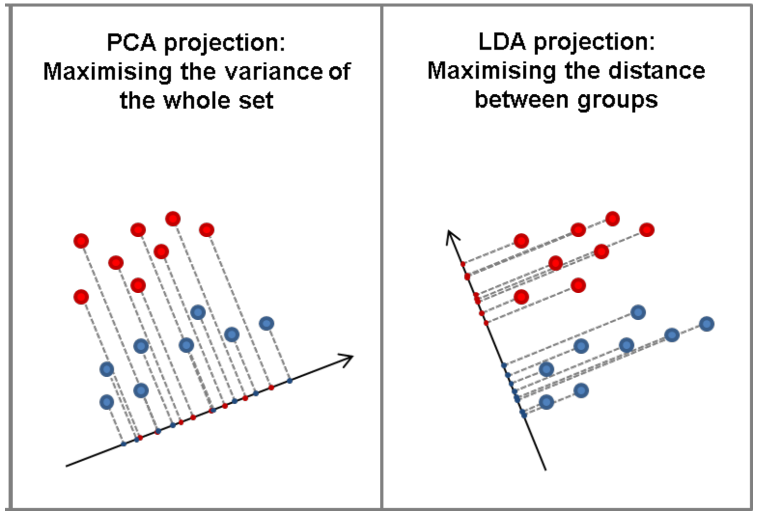
+ SMOTE



PREPROCESSING

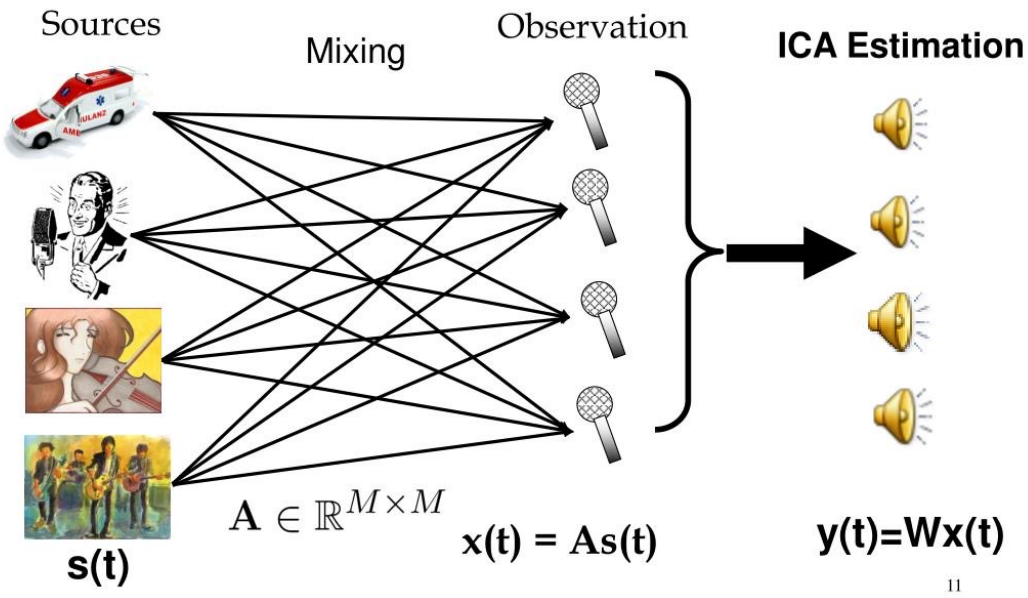
DIMENSIONALITY REDUCTION

+ Principal Component Analysis
+ Linear Discriminant Analysis



BLIND SOURCE SEPARATION

+ Independent Component Analysis



MODELS

ARCHITECTURES

+ Multiclass — output is a single correct prediction
+ Multilabel — labels are not mutually exclusive
+ Multitask — the same model generates two or more outputs

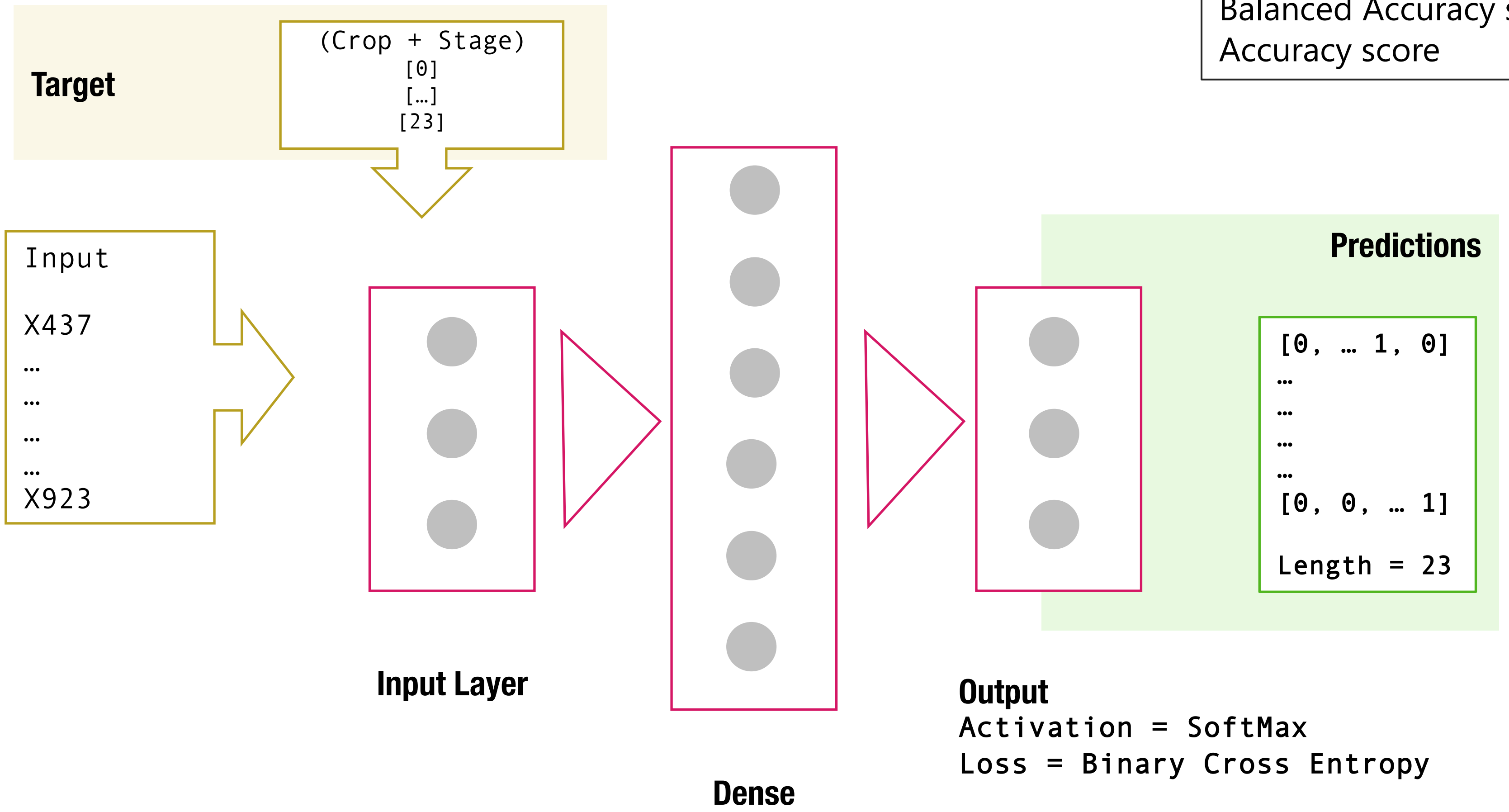
ALGORITHMS

+ Multi-layer Perceptron Classifier
+ Convolutional Neural Network
+ Chained Classifier (Gradient Boost)

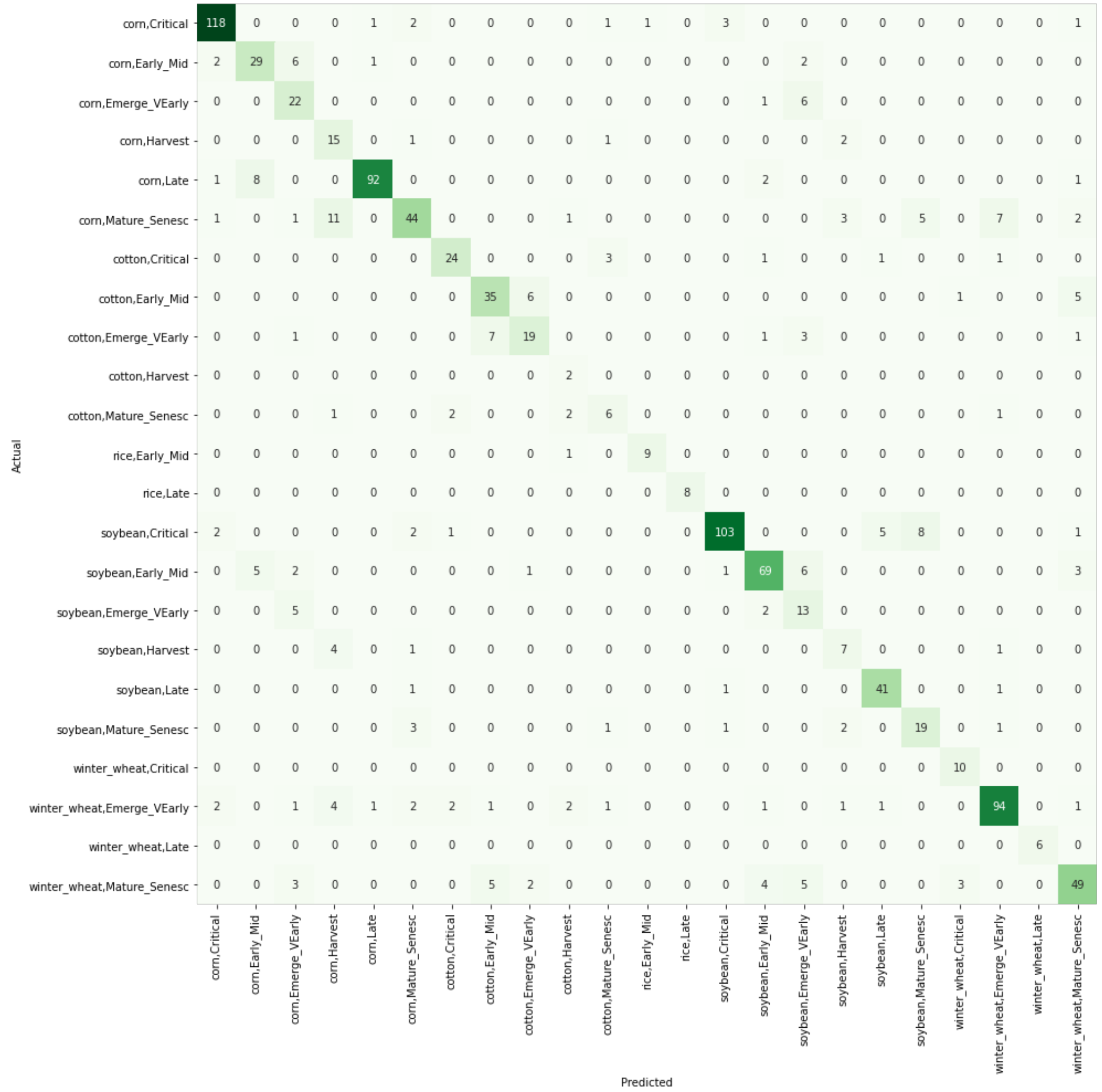
Multiclass Classifier

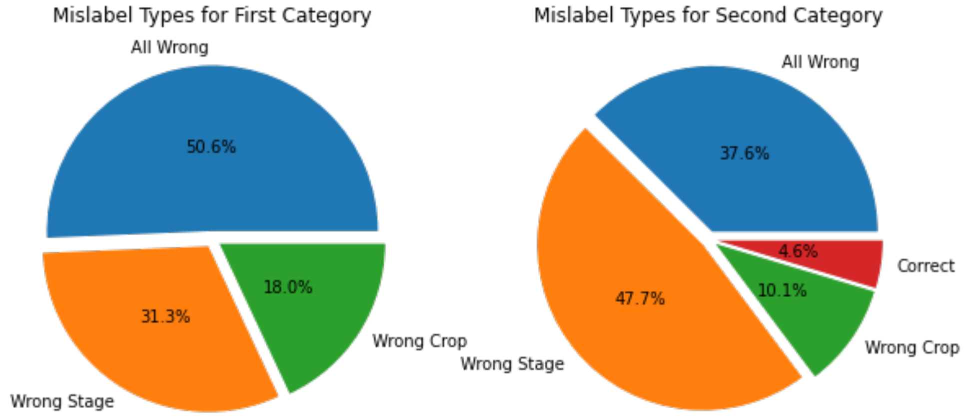
SCORES

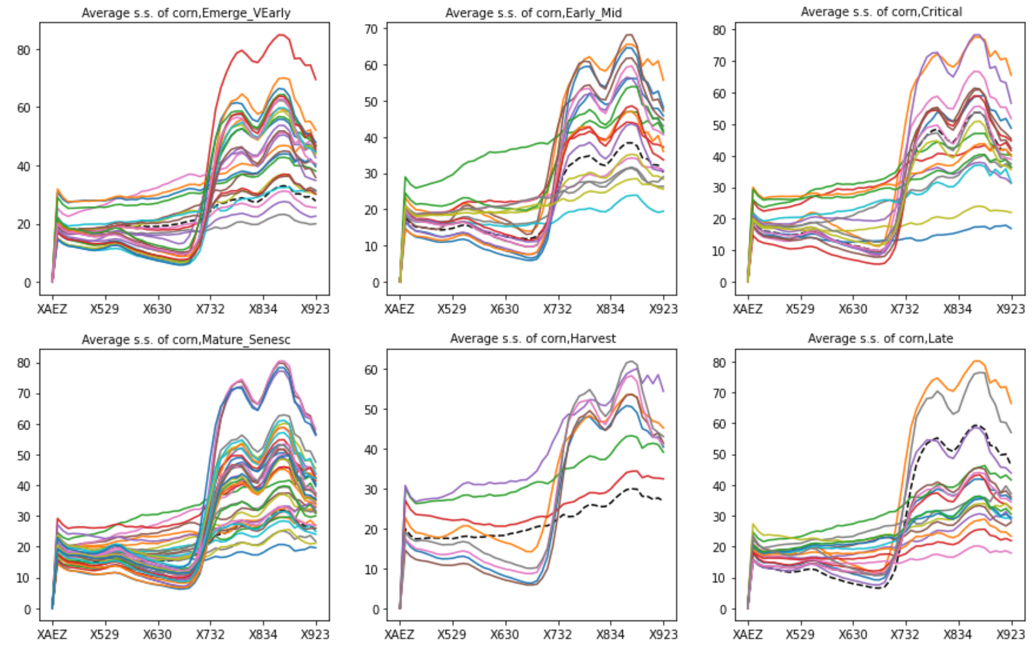
Balanced Accuracy score	79 %
Accuracy score	80 %



Multiclass Classifier



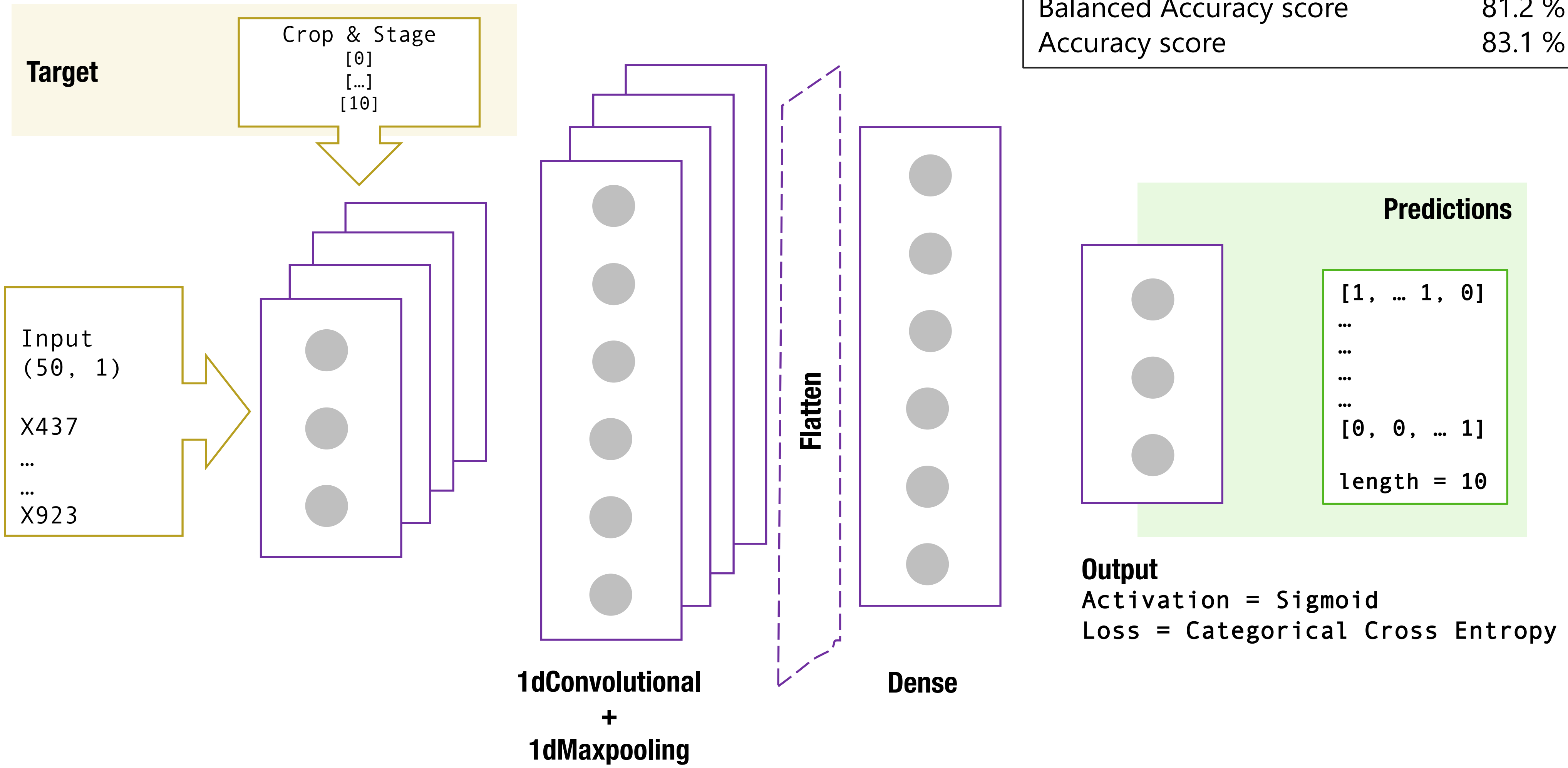




Multilabel Classifier

SCORES

Balanced Accuracy score	81.2 %
Accuracy score	83.1 %



Multilabel Classifier

Most common unidentified label combinations

('winter_wheat', 'Mature_Senesc')	38
('winter_wheat', 'Emerge_VEarly')	38
('soybean', 'Early_Mid')	38
('cotton', 'Early_Mid')	27
('corn', 'Mature_Senesc')	25
('soybean', 'Critical')	23
('corn', 'Emerge_VEarly')	22
('corn', 'Critical')	19
('soybean', 'Mature_Senesc')	18
('corn', 'Early_Mid')	18
('cotton', 'Critical')	17
('cotton', 'Emerge_VEarly')	16
('corn', 'Late')	15
('soybean', 'Emerge_VEarly')	14
('soybean', 'Late')	13
('cotton', 'Mature_Senesc')	11
('winter_wheat', 'Critical')	4
('winter_wheat', 'Late')	4
('rice', 'Early_Mid')	3
('rice', 'Late')	3

True labels	Predicted labels	
('winter_wheat', 'Mature_Senesc')	('cotton', 'Early_Mid')	6
	('cotton', 'Emerge_VEarly')	3
	('winter_wheat',)	3
	('winter_wheat', 'Critical')	3
	('Early_Mid',)	2
	('Mature_Senesc',)	2
	('corn', 'Emerge_VEarly')	2
	('cotton', 'Early_Mid', 'Mature_Senesc')	2
	('soybean',)	2
	('soybean', 'Early_Mid')	2
	('corn',)	1
	('corn', 'Mature_Senesc')	1
	('corn', 'winter_wheat', 'Mature_Senesc')	1
	('cotton',)	1
	('cotton', 'Mature_Senesc')	1
	('rice', 'Early_Mid')	1
	('soybean', 'Early_Mid', 'Mature_Senesc')	1
	('soybean', 'Mature_Senesc')	1
	('soybean', 'winter_wheat', 'Mature_Senesc')	1
	('winter_wheat', 'Early_Mid', 'Mature_Senesc')	1
	('winter_wheat', 'Emerge_VEarly')	1

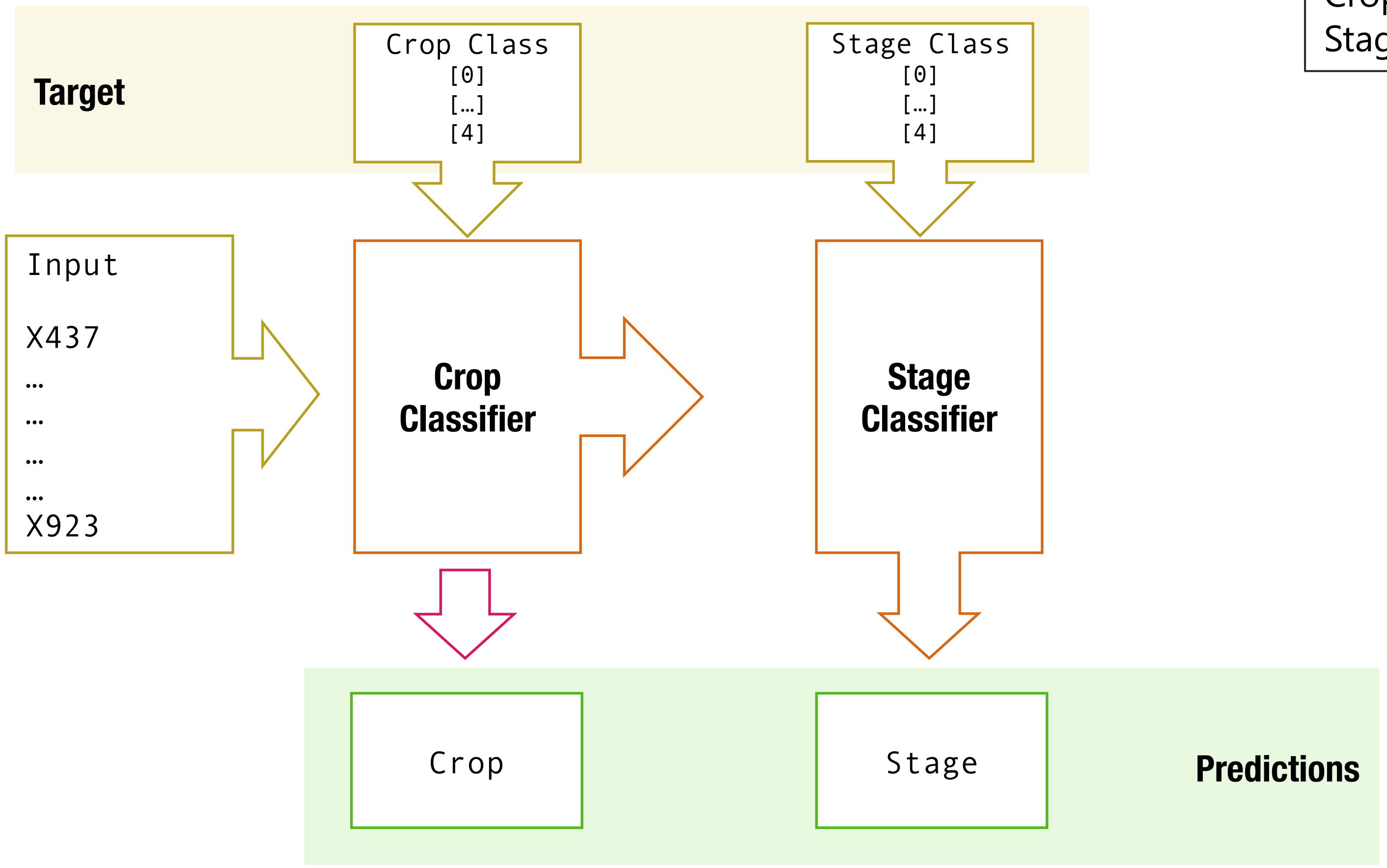
True labels	Predicted labels	
('cotton', 'Early_Mid')	('cotton', 'Emerge_VEarly')	15
	('winter_wheat', 'Mature_Senesc')	4
	('soybean', 'Early_Mid')	2
	()	1
	('Early_Mid',)	1
	('Early_Mid', 'Mature_Senesc')	1
	('cotton',)	1
	('cotton', 'Early_Mid', 'Emerge_VEarly')	1
	('winter_wheat',)	1

More wrongful, nonsystematic predictions omitted in the slide

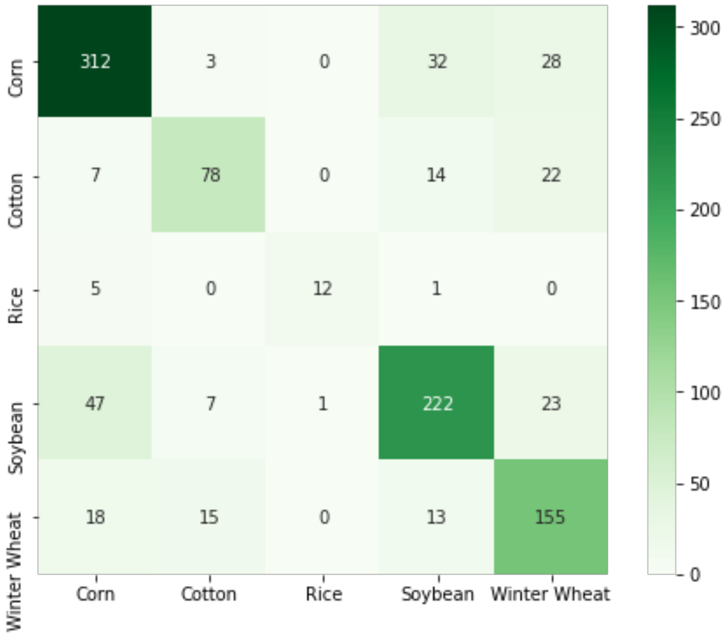
Multiclass Chained Classifier

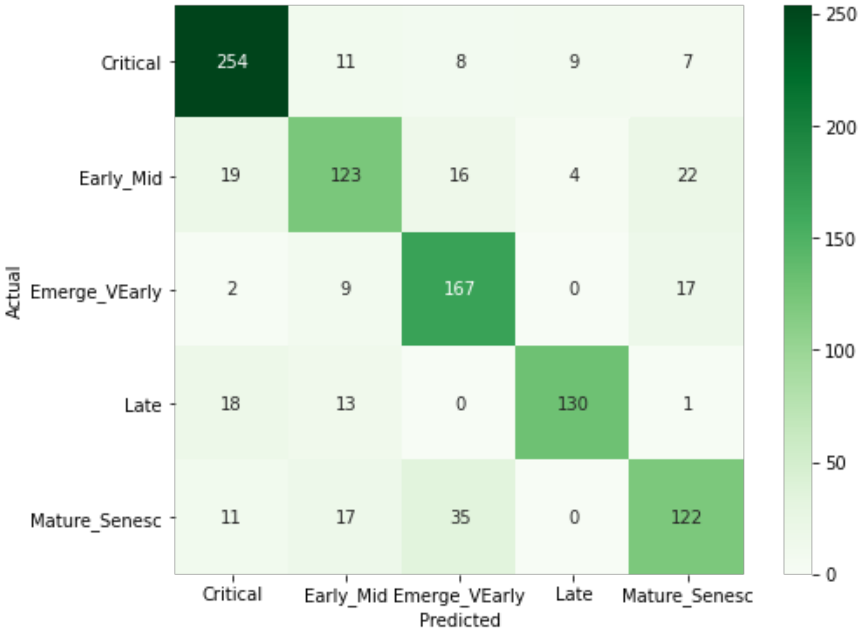
SCORES

Crop Accuracy score	76.06 %
Stage Accuracy score	79.1 %

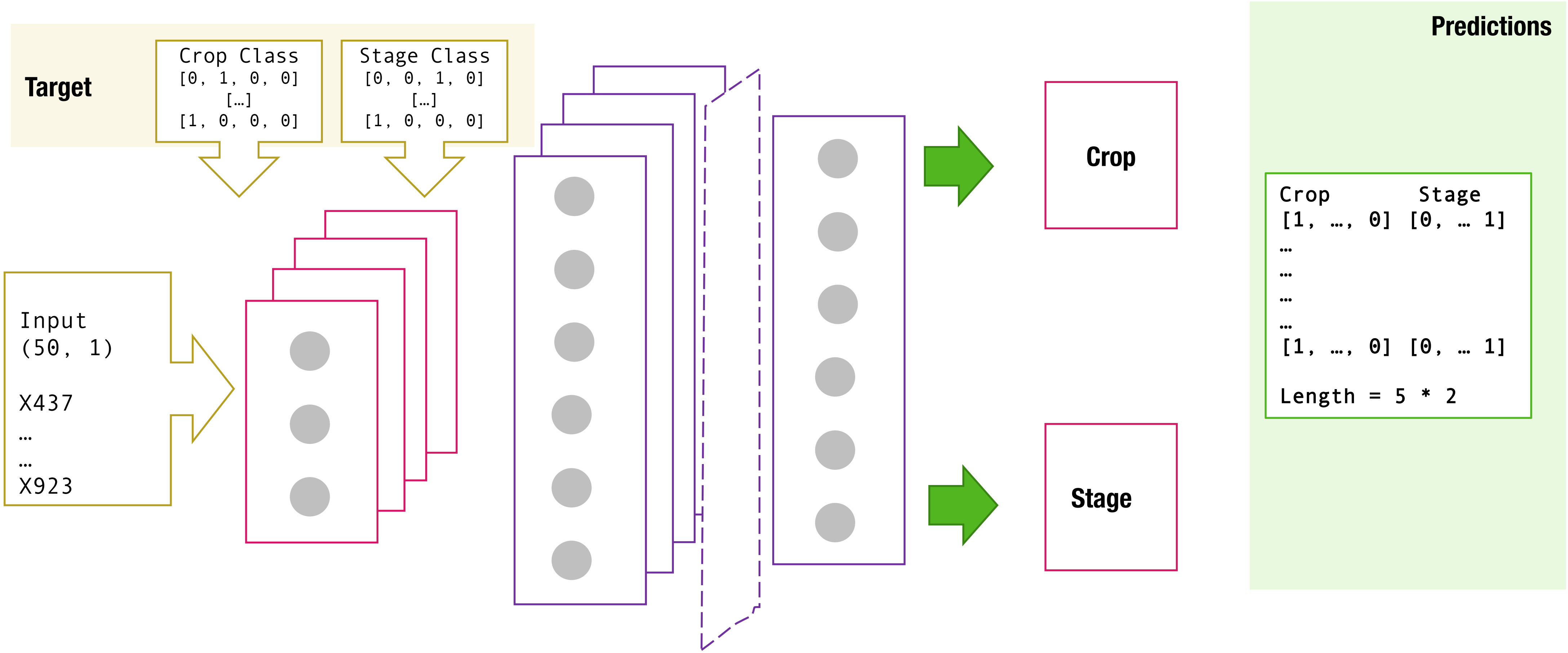


Multiclass Chained Classifier



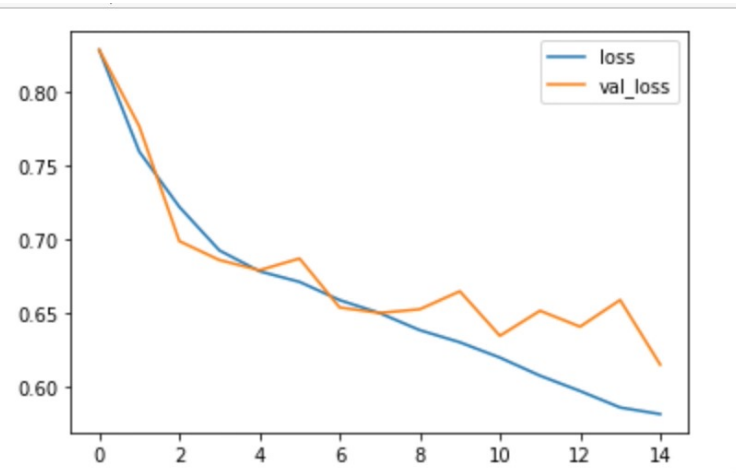


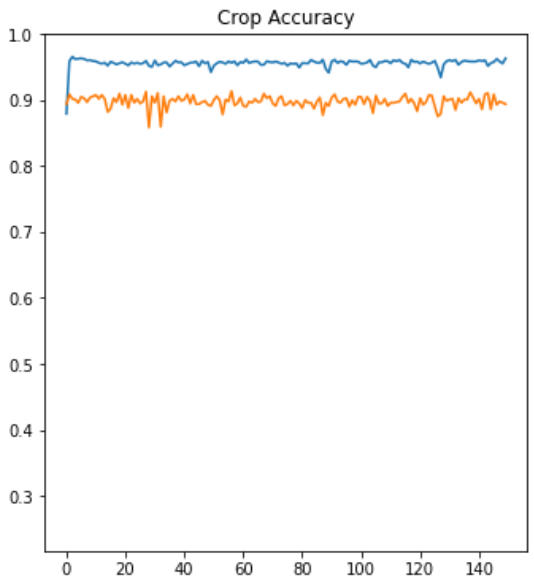
Multitask Classifier

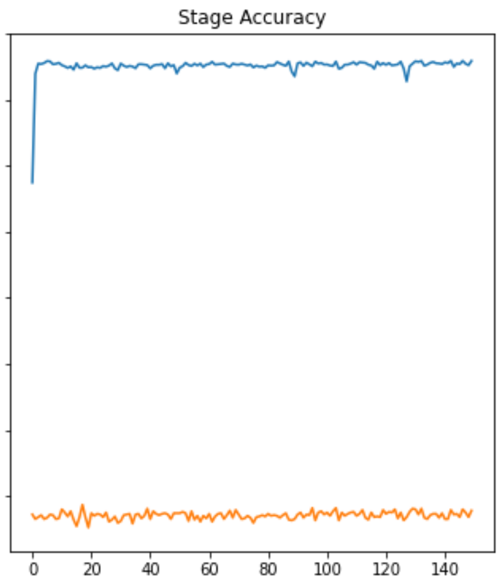


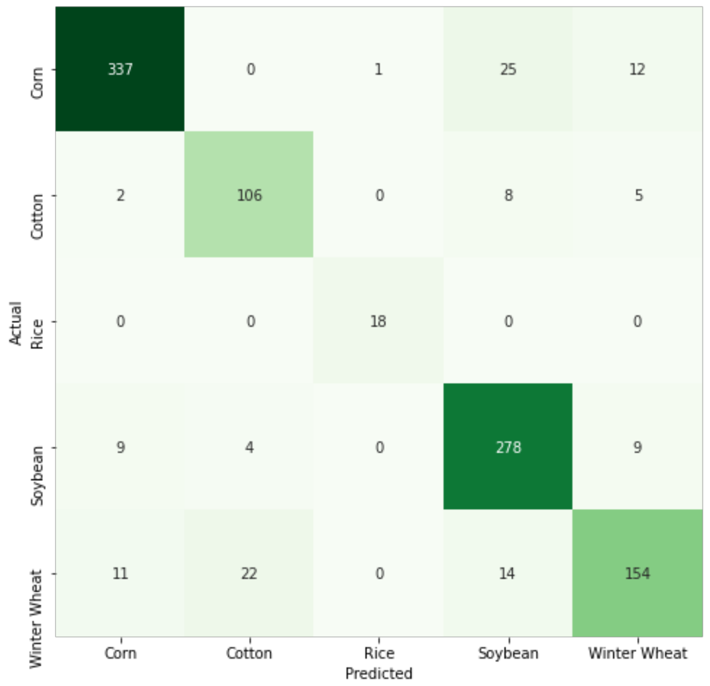
Multitask Classifier

+ Leveraging on stage as a tool to predict crops with more accuracy.
+ Upsample data highly reduces the performance of this model



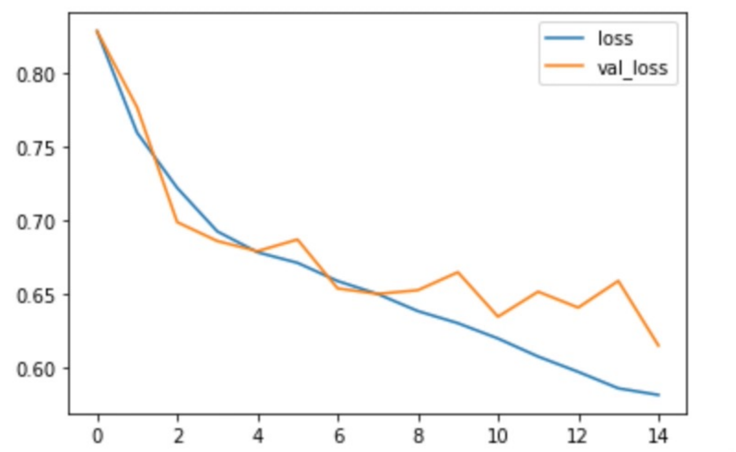


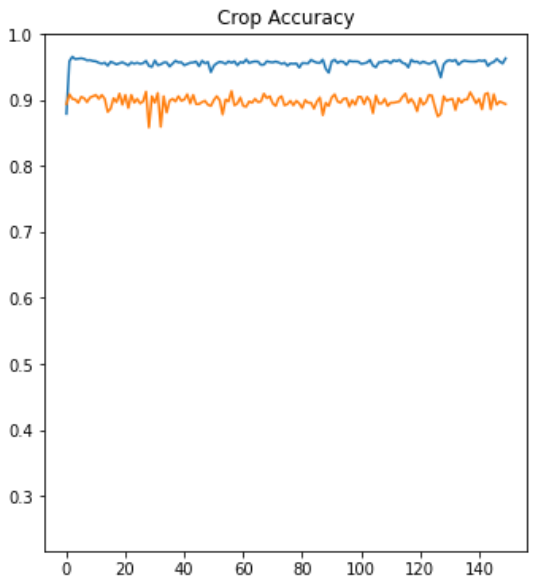


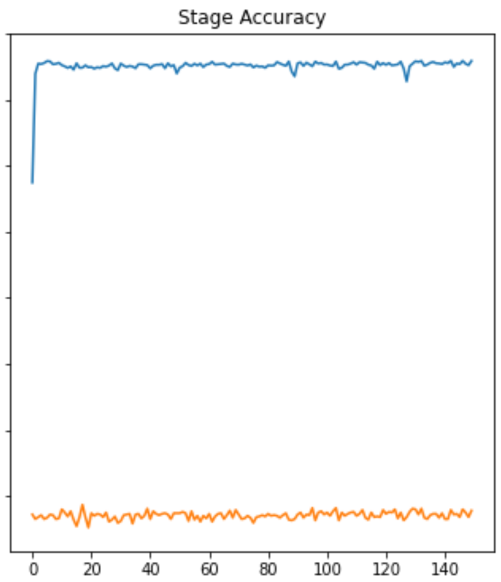


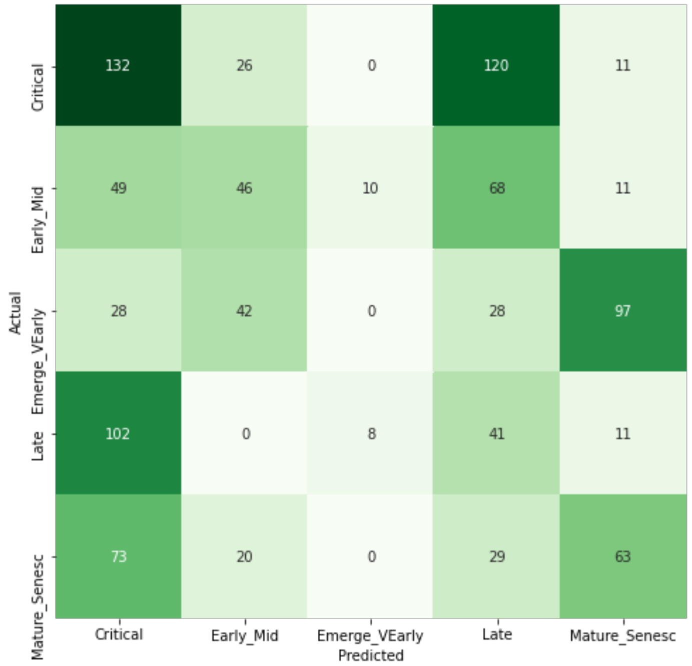
Multitask Classifier

+ Leveraging on stage as a tool to predict crops with more accuracy.
+ Up sample data highly reduces the performance of this model









CONCLUSION

- Multitask is a promising architecture
- ICA improves the results on fully connected neural networks, but it impacts negatively convolutional networks.
- Spectral data can be incredibly insightful even just using the first 50 bands.

NEXT STEPS

+ Pull larger data sample, train for results across geographic scope (Central Asia dataset available)

+ Pipeline of models, combining strength of different methods