AgSpace

Crop Type Classification

A Machine Learning Approach

William Manley

Machine Learning Summer Internship - AGSpace

AgSpace

Setting the Scene

- Identifying crop types from satellite imagery is hard from a human perspective - can a machine do better?

AgSpace

§1. Data Pipeline

- Data Retrieval

$\Downarrow$

- Data Filtering

$\Updownarrow$

- Data Processing

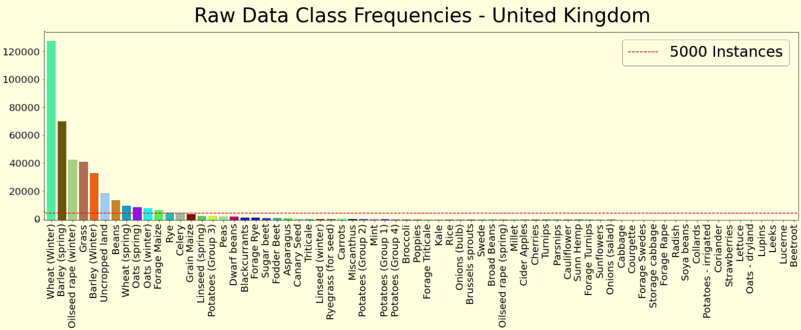
$\Downarrow$

- Data Exportation

AgSpace

Data Retrieval

- Unsure of the geographic dependency, we decided to focus exclusively on data from the UK.
- The raw, unfiltered data query retrieved a heavily skewed dataset, towards just a few crop types.
- A modest target of only 5000 instances per class can already only be achieved by a select few crop types.



AgSpace

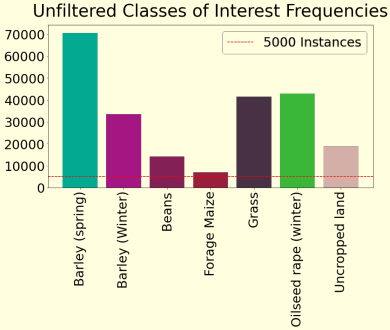
Data Filtering

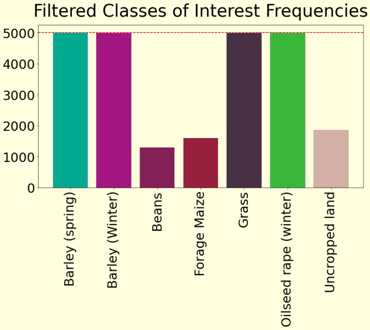
- Data must be within an *effective imagery period*. That is, for each crop type, we must define a lower bound for when the crop is first perceivable and an upper bound for it's expected date of harvest.
- Making sure that all field geometries are valid, contained in the raster, etc.
- There were many other important aspects to consider when filtering down our raw dataset to make the data fully effective.

AgSpace

Post-Filtering Analysis

- After filtering down to effective data, we are left with significantly less class counts.





AgSpace

Data Processing

- Individual fields need to be clipped from the raw satellite images.
- Images needs to be resampled to make it more manageable (RAM requirements, etc.).
- Individual Bands need to be combined to make useful features e.g. NDVI.
- The data needs to be standardized - normalization for ML performance and padding/cropping to square images.

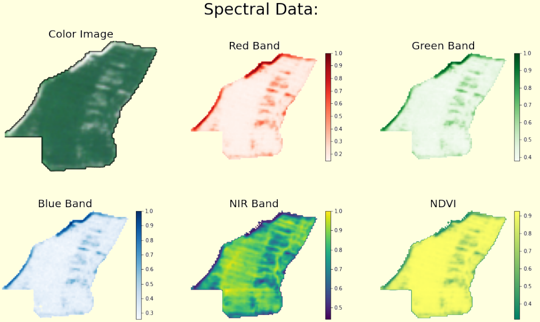
Extracting Individual Fields

- We first clip individual fields from the satellite images, using the geospatial raster and field polygon data.

AgSpace

Spectral Band Information

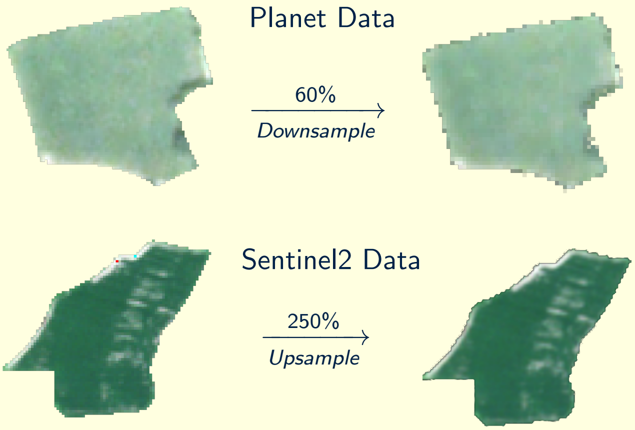
- Our images contain several spectral bands: red, green, blue and Near Infrared (NIR).
- From these bands we can calculate useful features such as the NDVI, which gives us a measure of the amount of photosynthesis and hence biomass within an image.



AgSpace

Resampling the Data

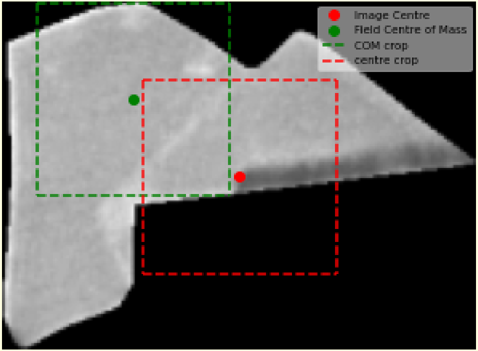
- Our images need to be resampled to make the data size more managable. Data is resampled to a common desired spatial resolution.



AgSpace

Data Standardization

- We normalize our pixel values to $[0, 1]$ to make them more suited to our ML algorithm.
- We also crop/pad images to a standardized dimension. Cropping by the fields 'centre of mass' achieved the best results.

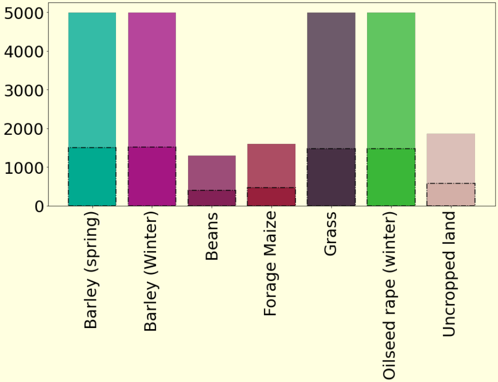


AgSpace

Data Exportation

- When making the test:train split, there are some important things to consider. For example, splitting so that training data contains few repeats of fields in the test set, to avoid overfitting.

30% Train:Test Split

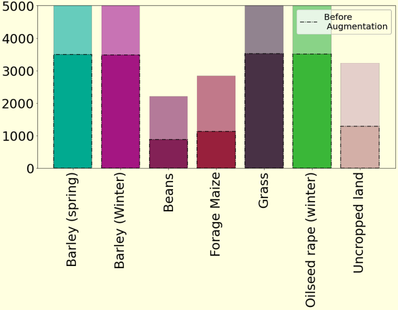


AgSpace

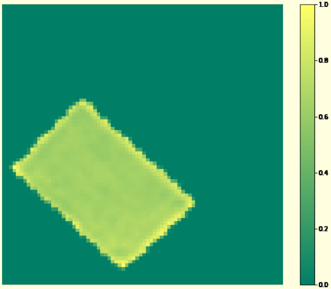
Data Augmentation

- To optimise training performance, we augmented some of our training set.
- Augmentation included random rigid body transformations and applying a small, spatially uniform, Gaussian noise.

Augmented Data



Augmented Field



AgSpace

§2. Timeseries Analysis

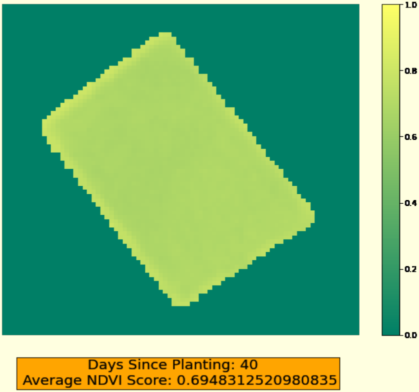
- It became apparent that the problem had a potentially strong time dependence.
- Crops appearance and features are dynamic and are constantly evolving throughout the growth cycle.
- A fully functional model will likely need be able to couple both the spatial and temporal aspects of the problem.

AgSpace

Evolution of a Sample Field

- An individual field is dynamic throughout it's growth period.

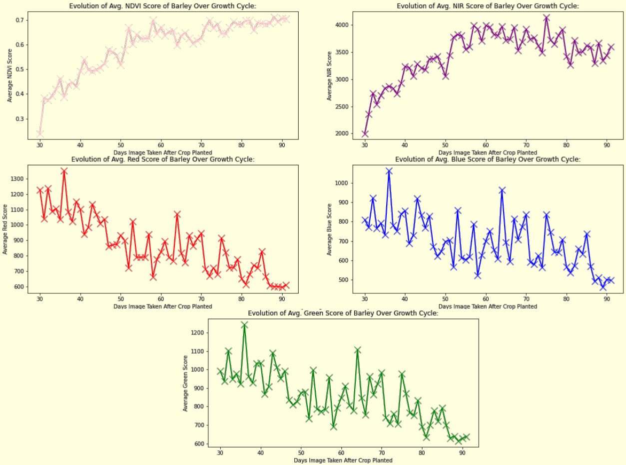
Barley NDVI Evolution



AgSpace

Crop Type Spectral Footprints

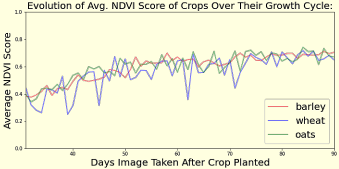
- There are some distinct correlations within a single crop type timeseries.

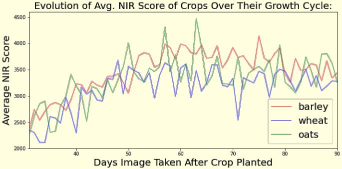


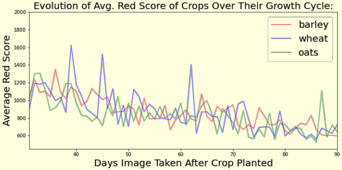
AgSpace

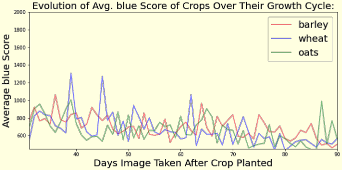
Comparing Crop Types

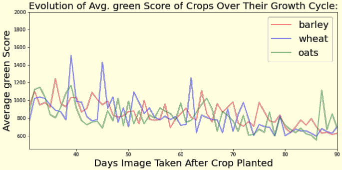
- When comparing crop types it seems that crops tend to follow a similar trend, with different levels of noisyness.











AgSpace

Some Final Thoughts

- Quantifying these results into features of a machine learning model is a difficult task.
- It is not immediately obvious how to couple the temporal aspect with the spatial nature of the problem.
- It is likely that seasonality effects on an annual basis could change the qualitative behaviour of the timeseries, although this is not something I explored.

AgSpace

Acknowledgements

- A massive thank you to everyone at AGSpace for providing me with this opportunity and for their continual support and guidance.
- In particular, thank you to Ivan Petrov for his supervision of our project and for providing a stimulating environment to learn in.
- Thank you to my direct collaborators Chang Liu and Dean Lundie for being a pleasure to work with and for all of their contributions to the project.