Guorun Huang, Project6 Report
Pattern Recognition and Computer Vision, 25Fall

Project Overview

In this project.

Background

CSIRO - Image2Biomass Prediction
Predict biomass using the provided pasture images。

Overview: Build models that predict pasture biomass from images, ground-truth measurements, and publicly available datasets. Farmers will use these models to determine when and how to graze their livestock.

Farmers often walk into a paddock and ask one question: “Is there enough grass here for the herd?” It sounds simple, but the answer is anything but. Pasture biomass - the amount of feed available - shapes when animals can graze, when fields need a break, and how to keep pastures productive season after season. Estimate incorrectly, and the land suffers; feed goes to waste, and animals struggle. Get it right and everyone wins: better animal welfare, more consistent production, and healthier soils. Current methods make this assessment more challenging than it could be. The old-school “clip and weigh” method is accurate but slow and impossible at scale. Plate meters and capacitance meters can provide quicker readings, but are unreliable in variable conditions. Remote sensing enables broad-scale monitoring, but it still requires manual validation and can’t separate biomass by species. This competition challenges you to bring greener solutions to the field: build a model that predicts pasture biomass from images, ground-truth measures, and publicly available datasets. You’ll work with a professionally annotated dataset covering Australian pastures across different seasons, regions, and species mixes, along with NDVI values to enhance your models.

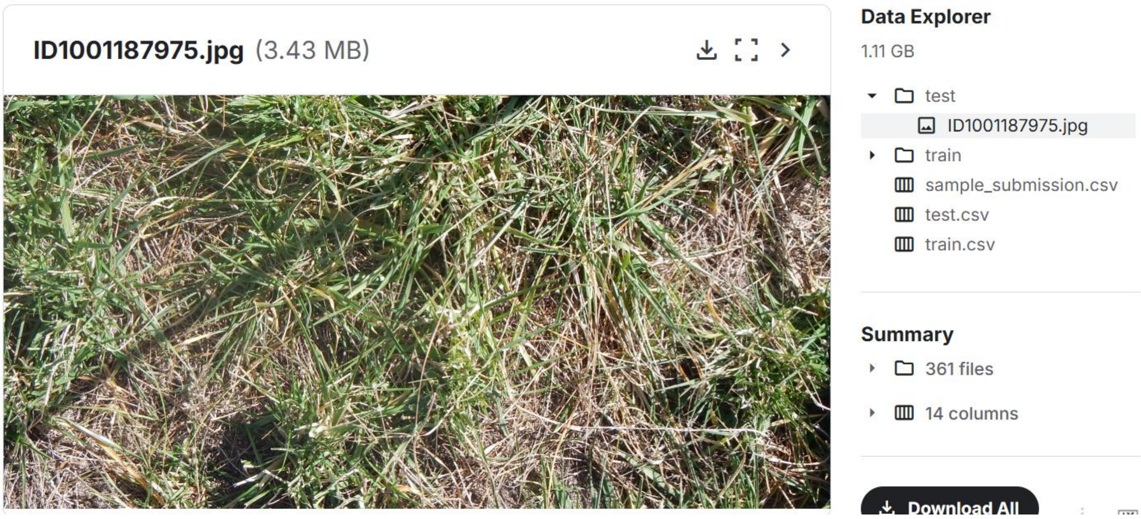
Evaluation-Scoring
The model performance is evaluated using a globally weighted coefficient of determination (R^2) computed over all (image, target) pairs together.
Each row is weighted according to its target type using the following weights:

Dry_Green_g: 0.1

Dry_Dead_g: 0.1
Dry_Clover_g: 0.1
GDM_g: 0.2
Dry_Total_g: 0.5
This means that instead of calculating R^2 separately for each target and then averaging, a single weighted R^2 is computed using all rows combined, with the above per-row weights applied.

Construct features from img

The dataset consists entirely of images showing grass growing on the ground with a gray- brown soil background. The grass varies in color, density, and condition—some patches are dry, while others are fresh.



The original dataset includes the following features:
sample_id — Unique identifier for each training sample (image).
image_path — Relative path to the training image (e.g., images/ID1098771283.jpg).
Sampling_Date — Date of sample collection.
State — Australian state where sample was collected.
Species — Pasture species present, ordered by biomass (underscore-separated).
Pre_GSHH_NDVI — Normalized Difference Vegetation Index (GreenSeeker) reading.
Height_Ave_cm — Average pasture height measured by falling plate (cm).
target_name — Biomass component name for this row (Dry_Green_g, Dry_Dead_g, Dry_Clover_g, GDM_g, or Dry_Total_g).
target — Ground-truth biomass value (grams) corresponding to target_name for this image.

To predict pasture biomass, I added the following features:

Color indices for separating green and dry grass:

Excess Green (ExG): highlights green vegetation. $ExG = 2G - R - B$ → Emphasizes green pixels and is useful for identifying fresh grass.

Excess Red (ExR) / ExGR: highlights yellow or brown plant material. $ExR = 1.4R - G$ → Enhances red components and helps detect dry or brown grass.

VARI (Visible Atmospherically Resistant Index): more stable than simple ratios. $VARI = (G - R) / (G + R - B)$ → Less sensitive to lighting changes and helps separate green vegetation from soil.

CIVE (Color Index of Vegetation Extraction): distinguishes green vegetation from bare soil or straw. $CIVE = 0.441R - 0.811G + 0.385B + 18.787$ → Designed to extract green vegetation and better separate grass from soil.

NDI (Normalized Difference Index, RGB version): $NDI = (G - R) / (G + R)$ → Similar idea to NDVI but based only on RGB; useful for distinguishing green and dry grass.

Histogram features:

RGB/HSV histograms: distribution and statistics (mean, variance, skewness) of each channel.

Hue distribution: calculates the proportion of green vs. brown/yellow hues, normalized by saturation to reduce soil interference.

RGB histograms and hue distributions both describe image color, but from different angles:

RGB histogram Counts pixel values in the R, G, and B channels.

Shows brightness distribution in each channel

Reveals overall color tendencies (e.g., greener or more reddish)

Sensitive to lighting changes Used for basic color analysis, such as estimating grass amount or overall color tone.

HSV histogram Based on Hue, Saturation, and Value:

H: color type

S: color purity

V: brightness More aligned with human perception and better for distinguishing green grass, dry grass, and soil. Less stable in low-saturation areas like shadows.

Hue distribution Converts the image to HSV or HSL and analyzes the hue component.

Hue reflects the actual color category

More robust to lighting and shadows

Good for separating green grass, dry grass, and soil Often used to estimate green- grass coverage.

If the goal is simply to separate green and dry grass, hue distribution is enough. If you also care about freshness (vivid green vs. dull yellow) and brightness, HSV histograms work better.

Texture and structure features:

GLCM/Haralick features: contrast, homogeneity, entropy. They are based on the Gray-Level Co-occurrence Matrix, which describes how pixel intensities relate to each other.

Dense green grass → finer, more uniform texture

Dry grass or straw → rougher texture, higher contrast

Soil → smoother texture

Coverage estimation:

Green coverage: the proportion of green pixels in the image.

How it’s calculated: Use ExG or another index (VARI, NDI) to score each pixel, set a threshold (e.g., ExG > 0 = green), then compute the percentage of green pixels.

Characteristics: It’s a summary metric — it only reflects overall coverage and doesn’t distinguish how strong the green color is. If we only use the mean ExG value, green coverage becomes almost redundant.

But if we use ExG distribution features (mean, variance, skewness, histogram), then green coverage becomes a simple and intuitive global feature that complements them.

So I keep:

ExG distribution features (mean, variance, histogram)

Green coverage as a lightweight global indicator of overall greenness

This way they are related but not repetitive.

Chose of features

After computing all the features, I combine them into a single vector so each image corresponds to one feature vector.

XGBoost can learn the weights by itself and decide which features matter more. The advantages are:

We don’t need to manually pick “the best” features — the model optimizes this on its own.

Combining different types of features makes the model more robust and prevents a single feature from failing under certain conditions.

However, not all features help the prediction. Some may be redundant or even add noise. By checking how much each feature contributes to the model, we can keep the important ones and remove the useless or repetitive ones, improving both efficiency and generalization.

Feature importance analysis Idea: during training, the model measures how much each feature reduces error or increases information gain at split nodes.

Outcome: the model produces a feature- importance score that shows which features are most useful for prediction.

In XGBoost, this can be accessed through model.feature_importances_.

Sentinel- 2 Data

There is no single pasture remote- sensing image dataset for Australia in 2015, but we can obtain the data from public satellite sources such as Landsat and Sentinel.

Landsat (NASA/USGS) Landsat provides 30 - meter multispectral imagery covering the entire world, including Australia. It is widely used for vegetation indices such as NDVI and EVI, and for monitoring pasture cover and biomass. For 2015, both Landsat 7 and Landsat 8 are available, but Landsat 7 suffers from striping issues, so Landsat 8 is generally preferred.

Sentinel- 2 (ESA Copernicus) Sentinel- 2 offers 10 - 20 m multispectral imagery. Sentinel- 2A began operating in July 2015, providing coverage over Australian pasture regions.

Key differences between Sentinel- 2 (S2) and Landsat 8 (L8):

Spatial resolution: S2: 10 m, 20 m, 60 m L8: 15 m, 30 m, 100 m → S2’s 10 m bands are better for smaller fields.

Revisit time: S2 (two satellites): 5 days L8: 16 days → S2 provides much more frequent observations.

Spectral bands: S2: 13 bands L8: 11 bands → S2 includes multiple red- edge bands, which are very useful for vegetation monitoring.

Because of these advantages, I chose Sentinel- 2 data.

The official documentation is here: https://knowledge.dea.ga.gov.au/notebooks/DEA_products/DEA_Sentinel2_Surface_Reflectance/

Processing level: The data is not raw Level- 1C imagery; it is Level- 2A Analysis Ready Data (ARD).

Data structure: You can think of it as a four- dimensional data cube:

X: longitude

Y: latitude

Time: a time series from 2015 to 2019 (roughly every 5 days, excluding cloudy scenes)

Bands: each pixel contains values for 10 spectral bands

All 10 m bands are resampled to 20 m so that every band has the same resolution.

DEA provides a Python environment (the Sandbox) where we load data using dc.load():

Product: ga_s2am_ard_3 (S2A) and ga_s2bm_ard_3 (S2B)

Measurements: the 10 bands used in the paper (nbart_blue, nbart_green, nbart_red, nbart_nir_1, etc.)

Location: the latitude–longitude bounding box of the study area

The output is an xarray.Dataset, essentially a large array containing reflectance values for all times and all bands.

Reproducing the paper's workflow
After loading the data, several steps are required:

1. Cloud masking The paper states that a cloud- detection algorithm removed more than 75% of cloudy pixels. DEA includes an oa_fmask band. Pixels marked as cloud or cloud- shadow must be set to NaN.

2. Resampling If using raw data, bands B2, B3, B4, and B8 are 10 m. They must be resampled to 20 m (nearest- neighbour) to align with the other bands.

3. Reflectance extraction The paper uses the median reflectance of all pixels inside each paddock. This requires a paddock boundary file (Shapefile/GeoJSON). We overlay the polygon on the satellite image, extract all pixels inside it, and compute the median for each band.

Approximating paddock locations
Because the dataset does not provide exact paddock coordinates, I estimated locations using the state- level coordinates with a small random offset. This is a rough approximation because a single state contains many paddocks in different conditions—some resting, some newly planted—and with different grass species and growth stages. Even paddocks close to each other can look very different, which is visible in the images.

In this project, my goal is mainly to enrich the information sources and explore this type of remote- sensing data. Unlike close- range camera images, satellite imagery is captured from far above and includes infrared bands, making it a very interesting data source.

Date formatting
DEA outputs time as an index in the format 2015- 07- 02 00:00:00. To match the image dataset, I convert it to the format 2015/7/2.

Determining the geographic window size
To decide how large the latitude – longitude window should be, I looked at the average size of Australian farms. Agricultural land covers about 55 – 61% of Australia (roughly 393 – 426 million hectares). There are about 136,000 farms, most of them livestock or pasture farms. This averages to about 3,100 hectares per farm, or 31 km².

Sentinel- 2 coverage: At around 30° S, each Sentinel- 2 scene covers roughly 290 km × 290 km. Spatial resolution depends on the band:

10 m (visible + NIR)

20 m (red- edge + SWIR)

60 m (atmospheric bands)

If we load a 1° × 1° area (~100 km × 100 km), then at 20 m resolution the image becomes a 5000 × 5000 pixel array. Loading a full month of data at this size can cause memory issues in the Sandbox, so I reduced the window to 0.1° × 0.1°.

Model Overview

Model 1: Images + ResNet

Each image is fed directly into a ResNet model, and the resulting embedding is used to predict the five target values.

Model 2: Hand- crafted features + metadata + XGBoost

All manually engineered features, together with metadata, are used as input to an XGBoost model.

Model 3: Hand- crafted features + metadata + MLP

The same set of engineered features and metadata is fed into a multilayer perceptron.

Model 4: Sentinel- 2 median band values + hand-crafted features + metadata + MLP

For each sample, the median value of all Sentinel- 2 bands is combined with the engineered features and metadata, then passed into an MLP.

Model 5: Fusion model (ResNet + XGBoost)

This model combines image embeddings from ResNet with predictions or features from XGBoost.

Kaggle Notebook

It’s an online interactive coding environment provided by Kaggle, with free access to CPUs and GPUs. It also integrates datasets, so you can directly load public datasets or competition data by selecting them through the “Add Data” button on the left side of the Notebook.

```
/kaggle/input/csiro-biomass/
     ├──── train.csv
     ├──── test.csv
     ├──── train/    ← train img
     └──── test/     ← test img
```

Results:

Model1

```
Epoch 1/5
Train Loss: 1318.909657, Val Loss: 710.542480
```

```
Target: Dry_Green_g - Train R^2: -1.0380 - Val R^2: -1.6912
Target: Dry_Dead_g - Train R^2: -0.9212 - Val R^2: -0.7906
Target: Dry_Clover_g - Train R^2: -0.3277 - Val R^2: -0.3036
Target: GDM_g - Train R^2: -1.7032 - Val R^2: -2.9436
Target: Dry_Total_g - Train R^2: -2.4547 - Val R^2: -3.2151
Saved best_model.pth

Epoch 2/5
Train Loss: 1423.471003, Val Loss: 617.834686
Target: Dry_Green_g - Train R^2: -0.8774 - Val R^2: -1.1764
Target: Dry_Dead_g - Train R^2: -0.7717 - Val R^2: -0.5498
Target: Dry_Clover_g - Train R^2: -0.2836 - Val R^2: -0.2160
Target: GDM_g - Train R^2: -1.4886 - Val R^2: -2.2375
Target: Dry_Total_g - Train R^2: -2.1233 - Val R^2: -2.3794
Saved best_model.pth

Epoch 3/5
Train Loss: 1007.043574, Val Loss: 547.849731
Target: Dry_Green_g - Train R^2: -0.6277 - Val R^2: -0.7864
Target: Dry_Dead_g - Train R^2: -0.5764 - Val R^2: -0.3335
Target: Dry_Clover_g - Train R^2: -0.2281 - Val R^2: -0.1017
Target: GDM_g - Train R^2: -1.1913 - Val R^2: -1.6664
Target: Dry_Total_g - Train R^2: -1.6770 - Val R^2: -1.7335
Saved best_model.pth

Epoch 4/5
Train Loss: 1022.889759, Val Loss: 584.445099
Target: Dry_Green_g - Train R^2: -0.3302 - Val R^2: -0.8993
Target: Dry_Dead_g - Train R^2: -0.4285 - Val R^2: -0.4330
Target: Dry_Clover_g - Train R^2: -0.1893 - Val R^2: -0.0476
Target: GDM_g - Train R^2: -0.8324 - Val R^2: -1.7476
Target: Dry_Total_g - Train R^2: -1.1787 - Val R^2: -1.9154

Epoch 5/5
Train Loss: 647.387251, Val Loss: 504.977234
Target: Dry_Green_g - Train R^2: 0.0005 - Val R^2: -0.5274
Target: Dry_Dead_g - Train R^2: -0.3272 - Val R^2: -0.2031
Target: Dry_Clover_g - Train R^2: -0.1636 - Val R^2: -0.0485
Target: GDM_g - Train R^2: -0.4491 - Val R^2: -1.2452
Target: Dry_Total_g - Train R^2: -0.6701 - Val R^2: -1.2970
Saved best_model.pth
Training completed.
Loaded best_model.pth

Train takes: 267.15 seconds

Prediction completed! Saved as submission.csv
                    sample_id     target
0  ID1001187975__Dry_Clover_g   2.251190
1    ID1001187975__Dry_Dead_g   6.236357
2   ID1001187975__Dry_Green_g   7.385307
3   ID1001187975__Dry_Total_g  13.002962
4         ID1001187975__GDM_g   6.946308

Total process takes: 267.59 seconds
```

Model2

```
Loading training data...
Training data shape: (1785, 9)
Target variables: ['Dry_Clover_g' 'Dry_Dead_g' 'Dry_Green_g' 'Dry_Total_g' 'GDM_g']

==================================================
Extracting features from unique images (will be reused for all targets)
==================================================
Number of unique images: 357
Extracting features...
Processing: 0/357
Processing: 100/357
Processing: 200/357
Processing: 300/357
Processing: 400/357
Processing: 500/357
Processing: 600/357
Processing: 700/357
Processing: 800/357
Processing: 900/357
Processing: 1000/357
Processing: 1100/357
Processing: 1200/357
Processing: 1300/357
```

```
Processing: 1400/357
Processing: 1500/357
Processing: 1600/357
Processing: 1700/357

==================================================
Training target: Dry_Green_g
==================================================
Training R²: 0.9997
Validation R²: 0.7613

All Feature Importance Scores:
                feature  importance
              Height_cm    0.195366
                  State    0.171623
             ExG_median    0.035714
             ExG_hist_9    0.033064
             ExG_hist_7    0.025618
             ExG_hist_8    0.025285
               ExG_mean    0.022831
             ExR_hist_2    0.019091
               B_hist_6    0.016018
           HSV_H_hist_2    0.015194
           HSV_H_hist_4    0.014647
               CIVE_std    0.014366
               B_hist_2    0.013824
           HSV_H_hist_0    0.013335
GLCM_dissimilarity_mean    0.013234
               R_hist_7    0.012498
              HSV_H_std    0.011571
                   NDVI    0.010780
           HSV_H_hist_1    0.009741
           HSV_H_hist_3    0.008043
             ExG_hist_4    0.007827
                ExG_max    0.007480
 GLCM_dissimilarity_std    0.007381
                Species    0.007322
           HSV_S_hist_3    0.007202
           HSV_V_hist_0    0.007017
               R_hist_4    0.006988
   GLCM_correlation_std    0.006584
          GLCM_ASM_mean    0.006310
               R_hist_2    0.006014
               G_hist_8    0.005977
             ExG_hist_5    0.005757
               G_hist_6    0.005737
           HSV_V_hist_2    0.005662
             ExG_hist_1    0.005465
           HSV_H_hist_6    0.005409
                 G_mean    0.005309
  GLCM_correlation_mean    0.005002
               R_hist_6    0.004994
                  Month    0.004917
               G_hist_1    0.004917
               R_hist_8    0.004813
             ExR_hist_4    0.004634
               ExR_mean    0.004433
           HSV_S_hist_2    0.004188
           HSV_H_hist_7    0.004188
              DayOfYear    0.003960
           HSV_V_hist_5    0.003846
               R_hist_5    0.003792
           HSV_H_hist_5    0.003685
              HSV_S_std    0.003632
       GLCM_energy_mean    0.003626
                  R_std    0.003621
  GLCM_homogeneity_mean    0.003615
     GLCM_contrast_mean    0.003597
              CIVE_mean    0.003577
             ExR_hist_1    0.003561
           HSV_S_median    0.003469
                ExG_min    0.003462
               G_hist_5    0.003368
             ExR_hist_6    0.003318
               B_hist_9    0.003205
             ExG_hist_3    0.003088
               R_hist_1    0.003055
      GLCM_contrast_std    0.003028
             ExR_hist_8    0.003025
             ExG_hist_0    0.002997
               B_hist_1    0.002949
             ExR_hist_0    0.002947
               G_hist_9    0.002926
               G_hist_4    0.002736
           HSV_V_hist_3    0.002726
             ExR_median    0.002707
           HSV_S_hist_1    0.002598
```

```
           ExG_hist_2    0.002545
         HSV_S_hist_0    0.002536
           ExR_hist_9    0.002517
             G_median    0.002516
             R_hist_3    0.002470
             R_hist_9    0.002464
             B_hist_0    0.002457
           HSV_H_mean    0.002447
             G_hist_0    0.002420
             B_hist_3    0.002407
             B_hist_8    0.002355
       green_coverage    0.002295
              ExG_std    0.002197
           ExR_hist_7    0.001976
            VARI_mean    0.001953
              ExR_std    0.001916
 GLCM_homogeneity_std    0.001781
             VARI_std    0.001750
           HSV_S_mean    0.001711
             B_hist_5    0.001704
             G_hist_2    0.001661
              NDI_std    0.001590
         HSV_V_hist_7    0.001489
             R_median    0.001483
           ExR_hist_3    0.001460
           ExG_hist_6    0.001458
               R_mean    0.001408
             B_hist_4    0.001238
               B_mean    0.001238
         HSV_S_hist_7    0.001099
                G_std    0.001035
             G_hist_7    0.001005
             B_median    0.000976
      GLCM_energy_std    0.000972
         HSV_V_hist_1    0.000967
                B_std    0.000905
         HSV_S_hist_6    0.000885
         HSV_H_median    0.000867
             B_hist_7    0.000852
         HSV_V_median    0.000843
           ExR_hist_5    0.000842
         HSV_S_hist_4    0.000784
             NDI_mean    0.000771
         HSV_V_hist_4    0.000738
         GLCM_ASM_std    0.000717
             G_hist_3    0.000712
             R_hist_0    0.000674
           HSV_V_mean    0.000434
         HSV_S_hist_5    0.000399
         HSV_V_hist_6    0.000363
            HSV_V_std    0.000230

==================================================
Training target: Dry_Dead_g
==================================================
Training R²: 0.9997
Validation R²: 0.5315

All Feature Importance Scores:
              feature  importance
           ExG_hist_0    0.061566
         HSV_H_hist_4    0.052781
                Month    0.046602
            HSV_S_std    0.034325
           ExR_hist_9    0.033688
             ExR_mean    0.030469
            Height_cm    0.025336
         HSV_H_hist_5    0.024713
           ExG_hist_1    0.021714
                State    0.020411
           HSV_H_mean    0.019679
           ExG_median    0.019110
    GLCM_contrast_std    0.018779
            DayOfYear    0.017329
           ExG_hist_7    0.017072
         HSV_H_hist_7    0.015259
             VARI_std    0.013377
              Species    0.012505
                 NDVI    0.012438
             G_hist_1    0.012304
           ExG_hist_3    0.011799
         HSV_H_hist_0    0.011178
            VARI_mean    0.011091
             G_hist_0    0.010686
             R_hist_9    0.010479
             B_hist_9    0.009820
         HSV_H_hist_2    0.009404
```

HSV_S_hist_0	0.009346
HSV_S_median	0.009098
ExG_hist_5	0.008851
GLCM_ASM_mean	0.008830
GLCM_ASM_std	0.008766
ExG_hist_2	0.008714
HSV_S_hist_2	0.008384
ExR_hist_0	0.008370
GLCM_energy_mean	0.008365
R_hist_2	0.007896
HSV_H_std	0.007850
B_hist_1	0.007765
HSV_V_hist_0	0.007746
ExR_std	0.007727
G_hist_6	0.007602
HSV_H_median	0.007400
HSV_S_mean	0.007359
GLCM_dissimilarity_std	0.007326
ExR_hist_8	0.007017
B_hist_5	0.006966
HSV_H_hist_3	0.006954
B_hist_8	0.006788
G_std	0.006103
R_std	0.005985
R_hist_1	0.005947
B_mean	0.005940
GLCM_correlation_std	0.005827
B_hist_0	0.005770
ExR_median	0.005756
R_hist_3	0.005714
GLCM_contrast_mean	0.005582
CIVE_mean	0.005515
B_hist_2	0.005155
G_hist_2	0.005139
GLCM_homogeneity_mean	0.005106
HSV_H_hist_1	0.005095
HSV_V_std	0.005037
B_median	0.005029
ExG_hist_9	0.004800
GLCM_correlation_mean	0.004675
G_median	0.004560
R_hist_6	0.004222
HSV_V_hist_6	0.004210
GLCM_energy_std	0.004164
HSV_H_hist_6	0.004108
R_hist_0	0.004084
CIVE_std	0.004077
HSV_V_mean	0.003990
green_coverage	0.003884
GLCM_homogeneity_std	0.003677
B_std	0.003645
HSV_V_hist_7	0.003450
R_mean	0.003305
ExR_hist_1	0.003213
HSV_S_hist_3	0.003153
ExG_mean	0.003130
ExG_min	0.003119
B_hist_7	0.003037
HSV_V_hist_3	0.003022
R_hist_7	0.003006
ExG_hist_4	0.002951
R_hist_5	0.002946
G_hist_9	0.002946
ExG_hist_8	0.002866
ExR_hist_4	0.002820
R_hist_8	0.002777
HSV_S_hist_4	0.002700
G_hist_4	0.002525
HSV_S_hist_7	0.002515
ExG_hist_6	0.002454
ExG_max	0.002425
NDI_mean	0.002377
B_hist_4	0.002355
R_hist_4	0.002339
ExR_hist_6	0.002289
ExG_std	0.002041
G_hist_7	0.002013
R_median	0.001936
G_hist_3	0.001829
HSV_V_hist_1	0.001813
B_hist_6	0.001773
NDI_std	0.001758
HSV_S_hist_6	0.001713
HSV_S_hist_5	0.001491
G_hist_8	0.001468
G_hist_5	0.001439
B_hist_3	0.001426

```
               G_mean    0.001413
             ExR_hist_2    0.001368
GLCM_dissimilarity_mean    0.001268
           HSV_V_median    0.001194
           HSV_S_hist_1    0.001086
             ExR_hist_7    0.000910
             ExR_hist_3    0.000749
             ExR_hist_5    0.000737
           HSV_V_hist_4    0.000399
           HSV_V_hist_5    0.000321
           HSV_V_hist_2    0.000307

==================================================
Training target: Dry_Clover_g
==================================================
Training R²: 0.9653
Validation R²: 0.7264

All Feature Importance Scores:
                feature  importance
             ExG_hist_7    0.062191
                Species    0.054106
           HSV_S_hist_4    0.041192
  GLCM_homogeneity_mean    0.038672
             ExR_hist_9    0.033686
                   NDVI    0.032954
                  State    0.032321
              CIVE_mean    0.029009
             ExR_hist_4    0.028560
             ExG_hist_5    0.023344
              VARI_mean    0.020606
           GLCM_ASM_std    0.019394
               G_hist_3    0.018814
           HSV_H_hist_7    0.018497
             ExG_hist_8    0.017124
           HSV_H_hist_5    0.016651
           HSV_H_hist_4    0.016042
                  Month    0.015373
           HSV_V_median    0.012567
               R_hist_8    0.011390
              DayOfYear    0.011170
               B_hist_9    0.011158
           HSV_H_hist_3    0.011115
         green_coverage    0.010774
           HSV_H_hist_1    0.010660
             ExG_hist_6    0.010610
             HSV_V_mean    0.010602
   GLCM_homogeneity_std    0.010415
              HSV_H_std    0.009975
               G_hist_2    0.009949
             ExG_hist_0    0.009801
             ExG_median    0.009656
                 G_mean    0.009612
               R_hist_1    0.009556
               G_median    0.009155
           HSV_S_hist_0    0.008902
   GLCM_correlation_std    0.008336
           HSV_V_hist_5    0.008022
           HSV_S_hist_1    0.007842
             HSV_H_mean    0.007618
           HSV_S_hist_6    0.007574
             ExR_hist_5    0.007306
               CIVE_std    0.007124
                  R_std    0.006856
             ExR_hist_3    0.006825
                ExG_min    0.006314
               NDI_mean    0.006128
           HSV_V_hist_1    0.005991
               R_hist_2    0.005838
             ExG_hist_9    0.005770
               B_hist_5    0.005707
 GLCM_dissimilarity_std    0.005404
                ExG_max    0.005320
               G_hist_6    0.005242
           HSV_H_hist_2    0.005187
               R_median    0.005130
              Height_cm    0.004887
             ExR_hist_7    0.004796
               R_hist_3    0.004752
           HSV_S_hist_3    0.004599
           HSV_V_hist_7    0.004574
               ExR_mean    0.004495
             ExR_median    0.004360
          GLCM_ASM_mean    0.004330
             ExR_hist_0    0.004241
           HSV_H_hist_0    0.004181
                  B_std    0.004150
```

GLCM_dissimilarity_mean	0.004124
B_mean	0.003855
ExG_mean	0.003815
G_hist_9	0.003722
HSV_V_hist_3	0.003704
HSV_S_mean	0.003581
VARI_std	0.003541
G_hist_1	0.003096
HSV_S_median	0.002999
GLCM_contrast_std	0.002816
HSV_V_std	0.002782
ExG_std	0.002535
B_hist_7	0.002449
HSV_V_hist_0	0.002441
G_std	0.002396
B_hist_3	0.002380
B_hist_2	0.002344
G_hist_0	0.002341
R_hist_9	0.002331
B_hist_0	0.002311
GLCM_contrast_mean	0.002298
B_hist_6	0.002292
ExR_hist_2	0.002226
ExR_std	0.002209
R_hist_7	0.002121
GLCM_energy_mean	0.002103
G_hist_5	0.002084
HSV_V_hist_2	0.002071
HSV_V_hist_4	0.002068
GLCM_energy_std	0.002033
HSV_S_hist_5	0.001970
HSV_H_hist_6	0.001928
B_hist_8	0.001897
R_hist_0	0.001689
GLCM_correlation_mean	0.001686
ExG_hist_1	0.001611
ExG_hist_3	0.001572
B_hist_1	0.001497
HSV_H_median	0.001420
G_hist_7	0.001413
B_median	0.001342
R_mean	0.001283
R_hist_6	0.001274
ExR_hist_6	0.001269
HSV_S_hist_7	0.001257
ExR_hist_1	0.001226
R_hist_4	0.001160
HSV_S_std	0.001131
R_hist_5	0.001052
NDI_std	0.000971
G_hist_8	0.000898
ExG_hist_2	0.000815
G_hist_4	0.000749
B_hist_4	0.000736
ExR_hist_8	0.000693
HSV_S_hist_2	0.000660
HSV_V_hist_6	0.000656
ExG_hist_4	0.000572

==
Training target: GDM_g
==
Training R^2: 0.9997
Validation R^2: 0.7140

All Feature Importance Scores:

feature	importance
Height_cm	0.168814
State	0.111645
HSV_H_hist_7	0.036288
NDVI	0.036227
GLCM_homogeneity_mean	0.035392
ExG_hist_9	0.035117
ExR_median	0.034051
ExG_hist_8	0.032511
ExR_mean	0.024822
ExR_hist_9	0.017990
GLCM_correlation_mean	0.016433
CIVE_std	0.016087
HSV_H_std	0.015391
Month	0.014624
ExG_median	0.014124
ExG_std	0.013343
ExG_hist_7	0.012998
CIVE_mean	0.012214
GLCM_ASM_mean	0.012032
DayOfYear	0.010972

B_hist_6	0.010497
NDI_mean	0.010243
HSV_H_hist_4	0.009205
ExG_mean	0.008776
HSV_H_hist_1	0.008576
G_hist_1	0.008496
R_hist_4	0.008303
HSV_H_hist_0	0.007789
ExG_hist_6	0.006942
GLCM_correlation_std	0.006855
Species	0.006766
R_hist_6	0.006438
R_hist_5	0.006292
green_coverage	0.006264
G_hist_0	0.006227
HSV_S_hist_7	0.006041
R_hist_1	0.006006
HSV_S_hist_6	0.005564
G_mean	0.005017
ExR_hist_8	0.004717
ExG_max	0.004687
HSV_V_hist_2	0.004645
ExR_hist_7	0.004620
HSV_V_hist_3	0.004447
R_hist_2	0.004035
HSV_H_mean	0.003962
HSV_S_hist_2	0.003735
GLCM_dissimilarity_std	0.003648
R_std	0.003613
G_hist_8	0.003530
ExR_hist_3	0.003494
G_hist_5	0.003388
B_hist_2	0.003351
ExR_hist_4	0.003320
HSV_V_hist_0	0.003167
GLCM_contrast_std	0.003162
G_median	0.003125
HSV_S_median	0.003072
ExR_std	0.003067
ExG_hist_5	0.002999
G_hist_4	0.002963
G_hist_2	0.002890
HSV_V_hist_5	0.002880
R_hist_0	0.002862
B_hist_9	0.002858
ExG_hist_1	0.002799
HSV_H_median	0.002786
VARI_mean	0.002773
GLCM_dissimilarity_mean	0.002727
B_hist_8	0.002718
HSV_V_hist_7	0.002671
GLCM_ASM_std	0.002649
HSV_S_std	0.002549
B_hist_3	0.002536
R_median	0.002527
B_std	0.002506
B_hist_1	0.002471
B_hist_7	0.002393
HSV_V_hist_6	0.002322
GLCM_energy_mean	0.002306
G_hist_9	0.002291
B_hist_0	0.002277
ExR_hist_6	0.002221
HSV_S_hist_0	0.002142
G_hist_7	0.002138
B_mean	0.002118
HSV_H_hist_3	0.001905
HSV_S_hist_1	0.001870
R_hist_3	0.001821
ExG_min	0.001817
G_std	0.001777
HSV_V_hist_1	0.001765
GLCM_contrast_mean	0.001667
ExG_hist_0	0.001659
GLCM_homogeneity_std	0.001640
HSV_H_hist_5	0.001606
HSV_V_std	0.001604
ExG_hist_2	0.001570
HSV_H_hist_6	0.001526
HSV_S_hist_4	0.001478
R_hist_7	0.001448
HSV_S_mean	0.001406
HSV_S_hist_3	0.001371
HSV_H_hist_2	0.001313
B_hist_5	0.001297
VARI_std	0.001292
ExR_hist_5	0.001088

```
             B_median    0.001039
              NDI_std    0.000960
               R_mean    0.000861
             HSV_V_mean    0.000811
             ExR_hist_1    0.000799
              B_hist_4    0.000739
           HSV_S_hist_5    0.000725
              R_hist_9    0.000715
              G_hist_6    0.000693
              R_hist_8    0.000643
             ExR_hist_0    0.000626
              G_hist_3    0.000581
         GLCM_energy_std    0.000555
             ExG_hist_4    0.000434
             ExR_hist_2    0.000430
           HSV_V_median    0.000413
             ExG_hist_3    0.000321
           HSV_V_hist_4    0.000209

==================================================
Training target: Dry_Total_g
==================================================
Training R²: 0.9998
Validation R²: 0.6618

All Feature Importance Scores:
                 feature  importance
               Height_cm    0.155413
                   State    0.142168
              ExG_hist_8    0.074864
            HSV_H_hist_7    0.034965
              ExG_hist_9    0.024906
    GLCM_correlation_std    0.020962
              ExG_median    0.020223
            HSV_H_hist_0    0.018592
               DayOfYear    0.016633
            HSV_H_hist_4    0.015985
               HSV_H_std    0.015385
                    NDVI    0.014663
                R_hist_2    0.012169
                   Month    0.012090
            HSV_S_median    0.011907
   GLCM_correlation_mean    0.010615
              ExG_hist_7    0.010396
              ExR_hist_8    0.010351
            HSV_V_median    0.009734
   GLCM_homogeneity_mean    0.009178
              HSV_H_mean    0.009085
                CIVE_std    0.008974
              ExR_hist_9    0.008930
                 Species    0.008734
                G_hist_5    0.008623
                 ExG_std    0.008013
                R_hist_4    0.007197
              ExG_hist_0    0.006992
            HSV_V_hist_2    0.006503
            HSV_S_hist_3    0.006292
                R_hist_6    0.006014
                ExG_mean    0.005753
            HSV_V_hist_5    0.005685
                G_hist_2    0.005390
            HSV_V_hist_0    0.005350
                G_hist_1    0.005348
                  G_mean    0.005278
            HSV_S_hist_7    0.005268
                B_hist_8    0.005221
            HSV_S_hist_6    0.005101
              HSV_S_mean    0.004990
                G_hist_8    0.004915
              ExR_hist_7    0.004893
                VARI_std    0.004830
            HSV_S_hist_2    0.004593
                G_hist_4    0.004466
                G_hist_3    0.004336
                R_hist_5    0.004268
                G_hist_6    0.004242
            HSV_H_median    0.004197
            HSV_H_hist_2    0.004169
              ExG_hist_3    0.004152
                R_hist_1    0.004146
              ExG_hist_2    0.004096
                G_hist_9    0.003950
                  B_mean    0.003933
                B_hist_6    0.003901
               HSV_S_std    0.003669
                 ExR_std    0.003658
                G_hist_7    0.003498
```

```
               VARI_mean    0.003461
         GLCM_energy_std    0.003393
              ExR_median    0.003370
                B_hist_1    0.003366
                   G_std    0.003350
                R_median    0.003226
              ExG_hist_4    0.003212
                 NDI_std    0.003196
        GLCM_energy_mean    0.003171
                B_hist_9    0.003161
                B_hist_0    0.003156
          green_coverage    0.003121
            HSV_V_hist_7    0.003120
                G_hist_0    0.003049
  GLCM_dissimilarity_std    0.003039
            HSV_H_hist_1    0.002958
            HSV_V_hist_1    0.002930
                R_hist_7    0.002906
              ExG_hist_5    0.002899
            HSV_H_hist_5    0.002785
            HSV_S_hist_1    0.002751
                   R_std    0.002682
            HSV_V_hist_3    0.002681
            GLCM_ASM_std    0.002665
    GLCM_homogeneity_std    0.002628
            HSV_H_hist_6    0.002602
      GLCM_contrast_mean    0.002596
                B_hist_7    0.002552
                ExR_mean    0.002298
                B_hist_4    0.002213
                R_hist_3    0.002163
                 ExG_max    0.002089
                B_hist_5    0.002066
                R_hist_0    0.001997
            HSV_S_hist_0    0.001991
 GLCM_dissimilarity_mean    0.001952
              ExG_hist_1    0.001861
                B_hist_3    0.001851
                G_median    0.001849
                B_median    0.001848
                R_hist_8    0.001769
               HSV_V_std    0.001756
                B_hist_2    0.001724
               CIVE_mean    0.001689
            HSV_V_hist_4    0.001588
              ExR_hist_4    0.001587
                 ExG_min    0.001569
                R_hist_9    0.001551
                NDI_mean    0.001512
           GLCM_ASM_mean    0.001511
              ExR_hist_6    0.001495
            HSV_H_hist_3    0.001439
            HSV_V_hist_6    0.001392
                   B_std    0.001327
       GLCM_contrast_std    0.001318
            HSV_S_hist_5    0.001270
            HSV_S_hist_4    0.001202
              ExR_hist_0    0.001157
              ExG_hist_6    0.001106
                  R_mean    0.001004
              ExR_hist_1    0.000814
              HSV_V_mean    0.000744
              ExR_hist_3    0.000732
              ExR_hist_5    0.000471
              ExR_hist_2    0.000220

Model training completed!

Train takes: 251.00 seconds
Loading test data...
Number of test images: 1
Extracting features...
Processing: 0/1
Adding 4 missing features: ['DayOfYear', 'Month', 'Species', 'State']...
Prediction progress: 0/5

Prediction completed! Submission file saved as submission.csv
Submission file shape: (5, 2)

Prediction statistics:
              count       mean  std        min        25%        50%  \
sample_id
Dry_Clover_g    1.0  10.611370  NaN  10.611370  10.611370  10.611370
Dry_Dead_g      1.0  17.626360  NaN  17.626360  17.626360  17.626360
Dry_Green_g     1.0  27.685587  NaN  27.685587  27.685587  27.685587
Dry_Total_g     1.0  34.636093  NaN  34.636093  34.636093  34.636093
GDM_g           1.0  37.497837  NaN  37.497837  37.497837  37.497837
```

```
                 75%        max
sample_id
Dry_Clover_g  10.611370  10.611370
Dry_Dead_g    17.626360  17.626360
Dry_Green_g   27.685587  27.685587
Dry_Total_g   34.636093  34.636093
GDM_g         37.497837  37.497837

Total process takes: 251.74 seconds
```

Model3

```
Using device: cpu
Loading training data...
Training data shape: (1785, 9)
Target variables: ['Dry_Clover_g' 'Dry_Dead_g' 'Dry_Green_g' 'Dry_Total_g' 'GDM_g']

==================================================
Extracting features from unique images (will be reused for all targets)
==================================================
Number of unique images: 357
Extracting features...
Processing: 0/357
Processing: 100/357
Processing: 200/357
Processing: 300/357
Processing: 400/357
Processing: 500/357
Processing: 600/357
Processing: 700/357
Processing: 800/357
Processing: 900/357
Processing: 1000/357
Processing: 1100/357
Processing: 1200/357
Processing: 1300/357
Processing: 1400/357
Processing: 1500/357
Processing: 1600/357
Processing: 1700/357
Total features extracted: 26

==================================================
Training target: Dry_Green_g
==================================================
Training MLP with 26 input features...
Early stopping at epoch 18
Training R²: 0.5408
Validation R²: 0.2815

==================================================
Training target: Dry_Dead_g
==================================================
Training MLP with 26 input features...
Early stopping at epoch 68
Training R²: 0.5807
Validation R²: 0.2652

==================================================
Training target: Dry_Clover_g
==================================================
Training MLP with 26 input features...
Early stopping at epoch 31
Training R²: 0.7043
Validation R²: 0.6625

==================================================
Training target: GDM_g
==================================================
```

```
Training MLP with 26 input features...
Early stopping at epoch 18
Training R²: 0.5519
Validation R²: 0.3956

==================================================
Training target: Dry_Total_g
==================================================
Training MLP with 26 input features...
Early stopping at epoch 23
Training R²: 0.4625
Validation R²: 0.1569

Model training completed!

Train takes: 112.91 seconds
Loading test data...
Number of test images: 1
Extracting features...
Processing: 0/1
Adding 3 missing features: ['Month', 'Species', 'State']...
Prediction progress: 0/5

Prediction completed! Submission file saved as submission.csv
Submission file shape: (5, 2)

Prediction statistics:
              count       mean  std        min        25%        50%  \
sample_id
Dry_Clover_g    1.0   0.000000  NaN   0.000000   0.000000   0.000000
Dry_Dead_g      1.0  10.643952  NaN  10.643952  10.643952  10.643952
Dry_Green_g     1.0  36.046967  NaN  36.046967  36.046967  36.046967
Dry_Total_g     1.0  62.635036  NaN  62.635036  62.635036  62.635036
GDM_g           1.0  43.762878  NaN  43.762878  43.762878  43.762878

                    75%        max
sample_id
Dry_Clover_g   0.000000   0.000000
Dry_Dead_g    10.643952  10.643952
Dry_Green_g   36.046967  36.046967
Dry_Total_g   62.635036  62.635036
GDM_g         43.762878  43.762878

Total process takes: 113.24 seconds
```

Model4

```
Number of unique images for feature extraction: 357

============================================================
Training target: Dry_Green_g
============================================================
Target: Dry_Green_g - Train R^2: 0.3517 - Val R^2: 0.3913

============================================================
Training target: Dry_Dead_g
============================================================
Target: Dry_Dead_g - Train R^2: 0.4560 - Val R^2: -0.0554

============================================================
Training target: Dry_Clover_g
============================================================
Target: Dry_Clover_g - Train R^2: 0.5832 - Val R^2: 0.6481

============================================================
Training target: GDM_g
============================================================
Target: GDM_g - Train R^2: 0.3524 - Val R^2: 0.5116

============================================================
Training target: Dry_Total_g
============================================================
Target: Dry_Total_g - Train R^2: 0.3898 - Val R^2: 0.2550

Model training completed!

Train takes: 126.72 seconds
Predicting: 0/5

Prediction completed! Submission file saved as submission.csv
Submission file shape: (5, 2)
```

```
Prediction statistics:
              count       mean  std        min        25%        50%  \
target_name
Dry_Clover_g    1.0   0.000000  NaN   0.000000   0.000000   0.000000
Dry_Dead_g      1.0   9.138971  NaN   9.138971   9.138971   9.138971
Dry_Green_g     1.0  35.033707  NaN  35.033707  35.033707  35.033707
Dry_Total_g     1.0  67.335342  NaN  67.335342  67.335342  67.335342
GDM_g           1.0  31.856451  NaN  31.856451  31.856451  31.856451

                    75%        max
target_name
Dry_Clover_g   0.000000   0.000000
Dry_Dead_g     9.138971   9.138971
Dry_Green_g   35.033707  35.033707
Dry_Total_g   67.335342  67.335342
GDM_g         31.856451  31.856451

Total process takes: 127.10 seconds
```

Model5

```
Target: Dry_Green_g - Train R2: 0.9814 - Val R2: 0.6464
   feature  importance
     State    0.075527
 Height_cm    0.043590
resnet_491    0.037221
resnet_280    0.032701
 resnet_66    0.022075
resnet_239    0.019123
resnet_123    0.018226
resnet_502    0.017673
resnet_381    0.016511
ExG_median    0.016320
resnet_268    0.016262
resnet_448    0.015967
resnet_344    0.014031
resnet_211    0.013761
resnet_187    0.013299
resnet_120    0.013206
resnet_326    0.012864
resnet_171    0.012679
 resnet_95    0.012222
resnet_391    0.011254
Target: Dry_Dead_g - Train R2: 0.9882 - Val R2: 0.4329
     feature  importance
  resnet_431    0.033122
  resnet_192    0.026251
  resnet_402    0.025850
  resnet_311    0.024245
  resnet_355    0.023618
HSV_H_hist_4    0.018974
  resnet_299    0.016562
  resnet_279    0.015421
  resnet_266    0.014394
  resnet_460    0.013055
  resnet_441    0.011611
  ExG_hist_0    0.011575
  resnet_507    0.011524
  resnet_264    0.010803
  resnet_397    0.010599
  resnet_187    0.010598
  resnet_490    0.010369
       Month    0.009865
  resnet_336    0.009060
  resnet_450    0.008796
Target: Dry_Clover_g - Train R2: 0.9995 - Val R2: 0.5878
     feature  importance
   resnet_66    0.073513
  resnet_107    0.060403
  resnet_498    0.055200
  resnet_341    0.037036
  resnet_149    0.031031
  resnet_118    0.022089
     Species    0.022080
  resnet_434    0.019981
  resnet_169    0.019733
  resnet_227    0.018198
   CIVE_mean    0.018043
  resnet_354    0.014940
  resnet_316    0.014265
```

```
HSV_H_hist_4     0.013963
   resnet_93     0.013567
  resnet_473     0.012549
  resnet_342     0.012456
  resnet_314     0.012339
       State     0.012172
   resnet_26     0.011669
Target: GDM_g - Train R2: 0.9511 - Val R2: 0.6518
     feature  importance
       State    0.063525
   Height_cm    0.041637
  resnet_491    0.037689
  resnet_381    0.018535
  resnet_344    0.018063
        NDVI    0.017855
   ExR_hist_9   0.015839
   resnet_66    0.015582
   resnet_94    0.013997
HSV_H_hist_4    0.012987
  resnet_277    0.012081
  ExR_median    0.011173
  resnet_171    0.010749
  resnet_211    0.010190
   resnet_55    0.010066
   resnet_71    0.009820
  resnet_270    0.009690
  resnet_168    0.009398
  resnet_327    0.008840
    ExR_mean    0.008576
Target: Dry_Total_g - Train R2: 0.9999 - Val R2: 0.5423
   feature  importance
     State    0.112873
 Height_cm    0.046597
resnet_208    0.038386
resnet_344    0.036757
resnet_346    0.028005
 resnet_71    0.022970
resnet_491    0.018806
resnet_302    0.014942
resnet_207    0.014513
resnet_381    0.012840
  resnet_8    0.012040
resnet_448    0.011862
resnet_480    0.011823
resnet_305    0.011377
 resnet_66    0.010643
      NDVI    0.009935
resnet_379    0.009445
resnet_249    0.009128
resnet_120    0.007926
resnet_426    0.007698
Model training completed.

Train takes: 182.58 seconds
Predicting: 0/5

Prediction completed! Submission file saved as submission.csv
Submission file shape: (5, 2)

Prediction statistics:
             count       mean  std        min        25%        50%  \
target_name
Dry_Clover_g   1.0  13.773639  NaN  13.773639  13.773639  13.773639
Dry_Dead_g     1.0  22.326494  NaN  22.326494  22.326494  22.326494
Dry_Green_g    1.0  17.789762  NaN  17.789762  17.789762  17.789762
Dry_Total_g    1.0  45.592762  NaN  45.592762  45.592762  45.592762
GDM_g          1.0  27.252375  NaN  27.252375  27.252375  27.252375

                   75%        max
target_name
Dry_Clover_g  13.773639  13.773639
Dry_Dead_g    22.326494  22.326494
Dry_Green_g   17.789762  17.789762
Dry_Total_g   45.592762  45.592762
GDM_g         27.252375  27.252375

Total process takes: 183.48 seconds
```


Acknowledgements

Thanks to Professor Maxwell and the TA Sihe for your support. I also referred

to the Kaggle, DEA, and various online tutorials throughout the project.