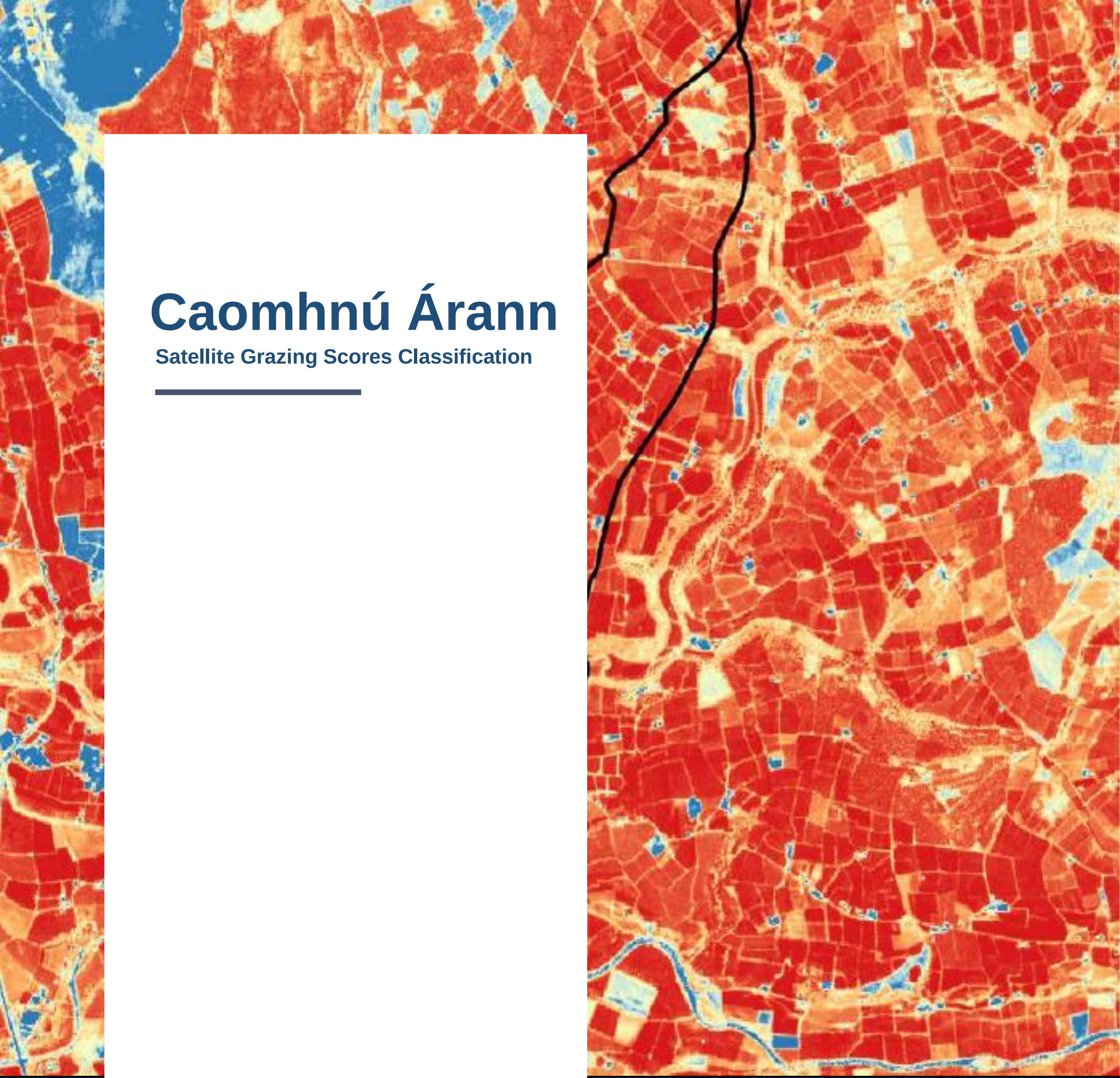


Caomhnú Árann

Satellite Grazing Scores Classification

APRIL 2019

Geo erospace
GEOSPATIAL AEROSPACE DATA COMPANY

Contents

1. Aim

The aim of this report is to asses the viability of using data captured from Sentinel 1 and Sentinel 2 to train machine learning models to predict grazing scores.

2. Machine Learning Predictions & Limitations

Before training any machine learning model, data exploration is vital to understand any limitations in the data being used. The break down of the data provided in the file "*grazing_scores_2016.shp*" is displayed in **Table 1** below.

Class	Number of Objects (Polygons)	Area of of Objects
1	10	0.02 km^2
2	67	0.16 km^2
3a	86	0.73 km^2
3b	202	0.87 km^2
4	253	2.33 km^2
5	400	5.13 km^2

Table 1: *Grazing_scores_2016 class breakdown*

The application of most machine learning approaches requires large volumes of data for accurate models to be generated. Unfortunately, with the low number of individual objects (1018), training an accurate model would be quite difficult. It would also be recommended to ignore class "*1*" until more data is acquired. Data used for training a machine learning model should ideally have a relatively equal number of objects in each model. If not approximately equal, a trained model will be predisposed to classify features to a specific class as it is encountered more often during its training process. Training a model on the data above would require only using a subset of classes "*3b*","*4*" and "*5*" so that the number of objects used in training the model is closer to classes "*2*" and "*3a*".

3. Viability of Satellite Data for Classification

To assess the viability of using satellite data to classify land into grazing scores, two sources of satellite data were explored, Sentinel 1 and Sentinel 2. Only satellite data collected in 2016 was used in this assessment so that the data collected would coincide with the date the grazing scores were obtained.

Sentinel 1 is a Synthetic Aperture Radar (SAR) satellite. It is capable of collecting data through cloud cover and even at night. During 2016 two Sentinel 1 satellites were operational SENTINEL 1A and SENTINEL 1B. SENTINEL 1B only became active mid-2016. The data used for this assessment was in the "*Ground Range Detected*" (GRD) format with a resolution of 10m2.

Sentinel 1 GRD data provides data in two configurations VV and VH. Due to its cloud penetrating capabilities, 112 Sentinel 1 images were captured over the Aran Islands. For this assessment, only 12 images were used, roughly one each month.

Sentinel 2 is multispectral satellite. It has 13 spectral bands. Four of these bands have a resolution of 10m2, band 2 (blue), band 3(green), band 4(red) and band 8(NIR). Due to their higher resolution, only theses bands were used during this assessment. As Sentinel 2 cannot penetrate cloud cover and as only one of the two Sentinel 2 satellites were available in 2016. Only 3 Sentinel 2 predominantly cloud free images could be found in 2016 and each of which has some cloud cover over at least one of the islands. Data use was atmospherically corrected L2A data.

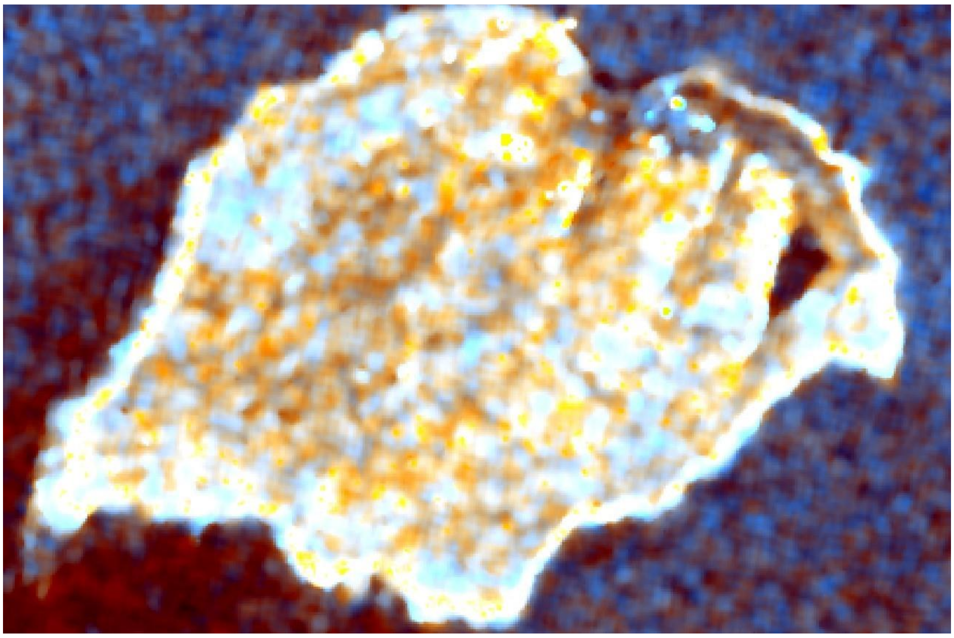


***Figure 1**: Sentinel 1 28/02/2016 captured over Inisheer*

Figure 2: Sentinel 2 29/05/2016 captured over Inisheer.

For this assessment, data was extracted from the Sentinel images as an average value within each polygon in the labelled data. **Figure 3 and 4**, illustrates the central tendency from the NDVI and VV/VH ratio date the data was captured. In both figures, the vertical dashed lines indicate the date that the data was captured. Class 1 is excluded from theses plots due to the low number of samples and high variance in the data captured.

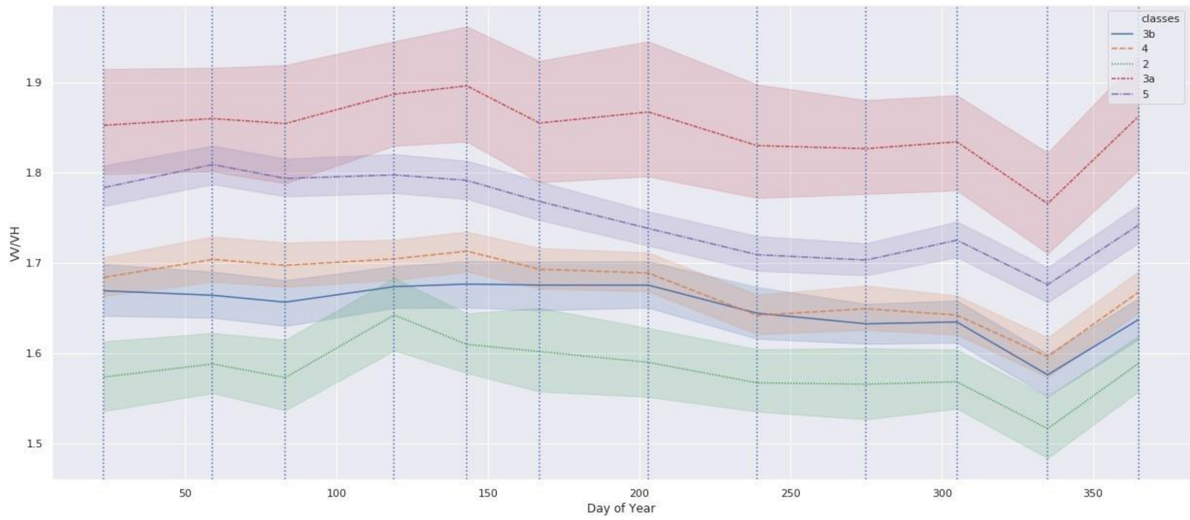


Figure 3: *Plot of VV/VH from Sentinel 1 in 2016*

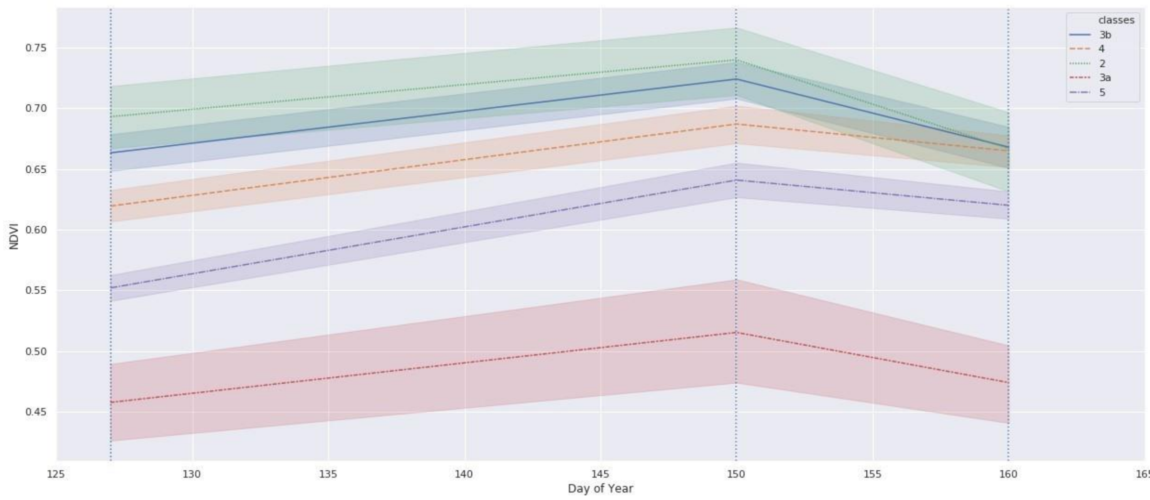


Figure 4: *Plot of NDVI from Sentinel 2 in 2016*

4. Results

Although at first, it appears that using the spectral response from the surface it should be possible to classify the areas into their different grazing scores. Further examination of the data reveals that there is a high degree of variation in the spectral response in the data as can be seen in the box plots of data from individual days as displayed in **Figures 5 and 6**.

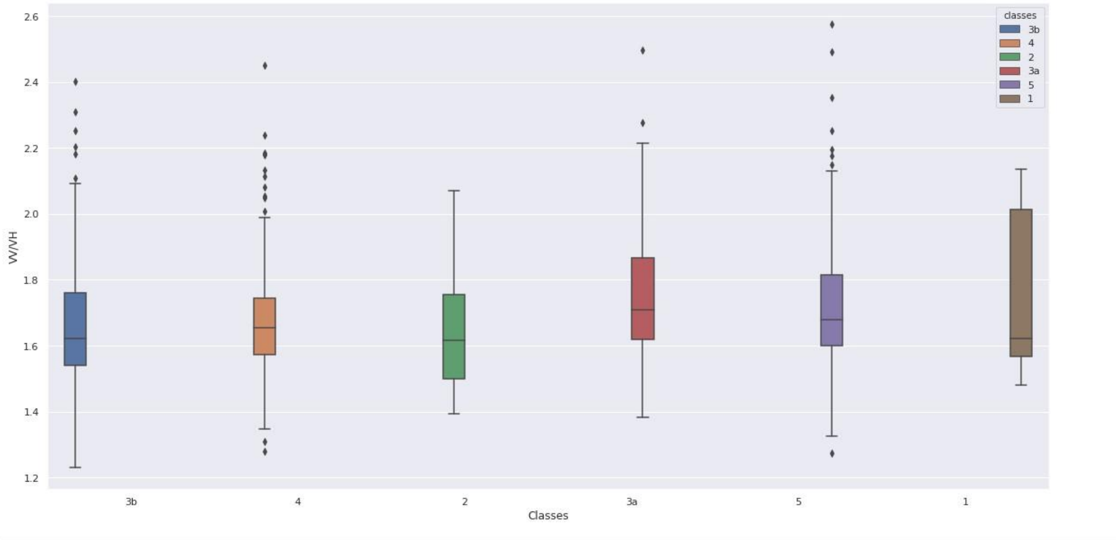


__Figure 5__: Plot of VV/VH from Sentinel 1 26/08/2016

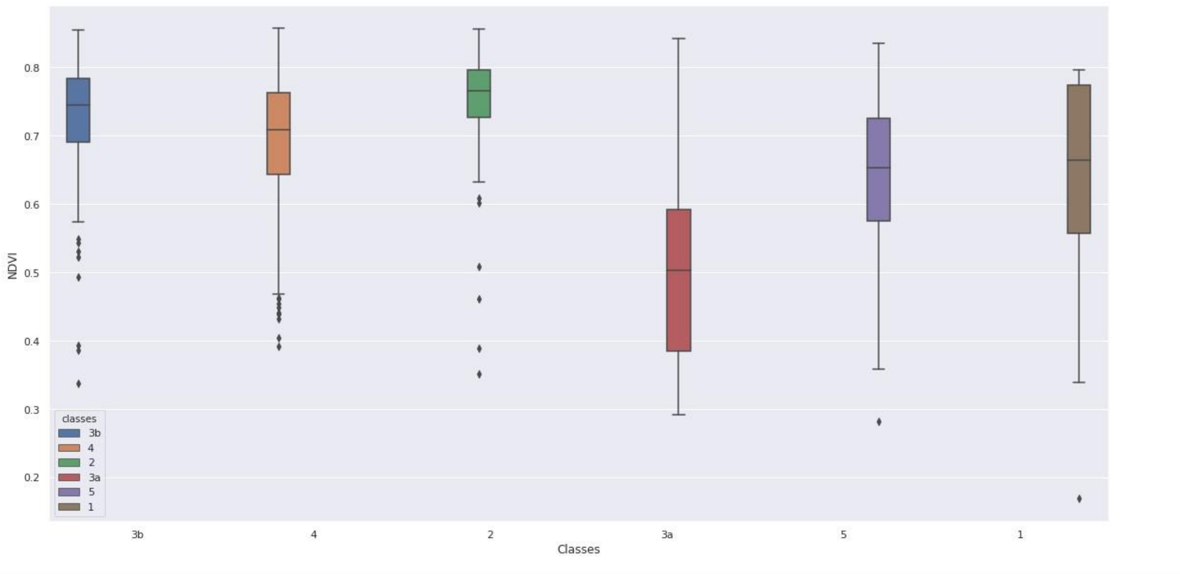


Figure 6: Plot of NDVI from Sentinel 2 29/05/2016

With large overlapping value ranges and a large number of outliers in each class, classification of the individual grazing scores was not possible.

The classification was attempted using three different Machine Learning techniques, Random Forests, Support Vector Machines and Multi-layer Perceptron. Unfortunately, no model could be trained that could reliably classify the different land cover types using the spectral response from the two satellite sources.

5. Recommendations

Although in this instance it was not possible to successfully train a model which could classify the areas into their respective habitats, analysis of the data does suggest that with modifications satisfactory results could be achieved. Despite there was a large degree of variation in the spectral response over each habitat type, **Figures 3 and 4** do indicate that there is a difference in the central tendency of the data for each habitat type. Following this, five recommendations would be advised for any future study.

1. Classify each field separately

The workflow undertaken in this assessment used the provided vector data to extract the mean spectral response of the object. However, as can be seen in figure 7, many polygons cross several fields each of which may have a separate spectral response. The polygons also include the boundary walls between fields which could be a cause of the high degree of variance in the data.

Figure 7: Inisheer habitat polygons with high resolution aerial imagery

2. Higher Resolution data

While Sentinel 1 and 2 are an excellent source of free data, due to their lower (10m) resolution, the relatively small field sizes on the Aran Islands require higher resolution data to achieve satisfactory results.

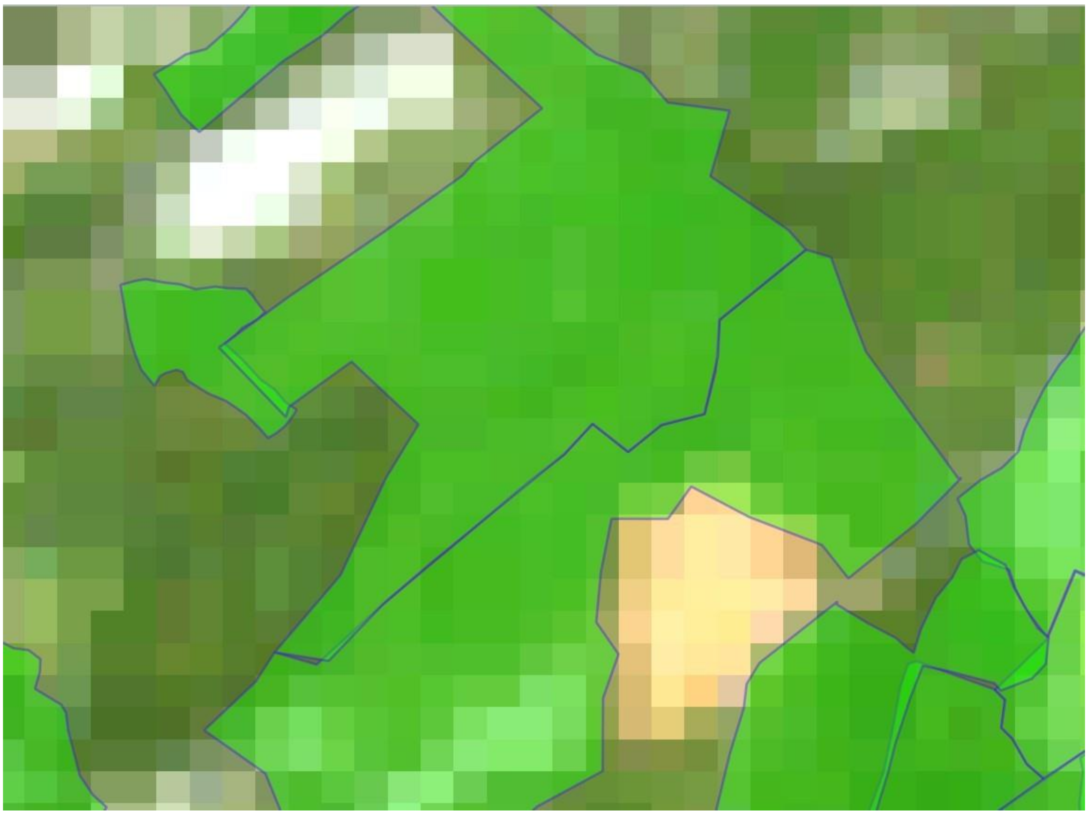


Figure 8: Inisheer habitat polygons with Sentinel 2 10m resolution imagery

3. Pre-classification/ Stacked Classification

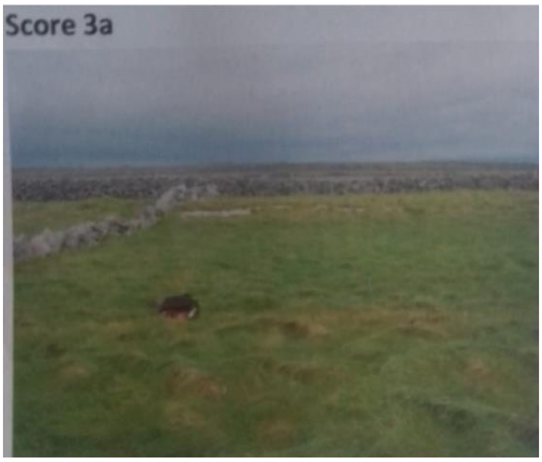
Due to the complexity of the land cover on the island and the low number of training examples, applying a series of simpler models may produce better results.

An example of how this approach could work on high-resolution images is to have three classification steps.

- Classify field boundaries
- Classify vegetated and non-vegetated regions.
- Classify the habitat of the field based on the spectral response of the vegetation and the percentage of bare ground.

4. Divide habitat 3a into two classes during the training

Habitat 3a contains two types of land cover areas which have dramatically different spectral responses. This is likely the reason why habitat 3a has one of greater degrees of variation as visible in **Figures 5 and 6**.



5. Obtain more labelled data

Due to the nature of training machine learning models, higher quantities of labelled data is required. This could partially be achieved through breaking up the original habitat labels into separate fields. However, ideally, new data would need to be acquired. The open street map data set may be useful data for obtaining polygon segments of fields on the islands as it appears that much work has been already been done digitising field boundaries. However, significant time would be needed to check the quality of the digitisation.