EDA Report

Contents

Chapter 1

Dataset Overview

The EuroSAT RGB dataset consists of 10 balanced land–use land–cover classes. Each image is of shape $\mathbf{64 \times 64 \times 3}$ (RGB). A summary of class distribution is shown below.

1.1 Class Distribution

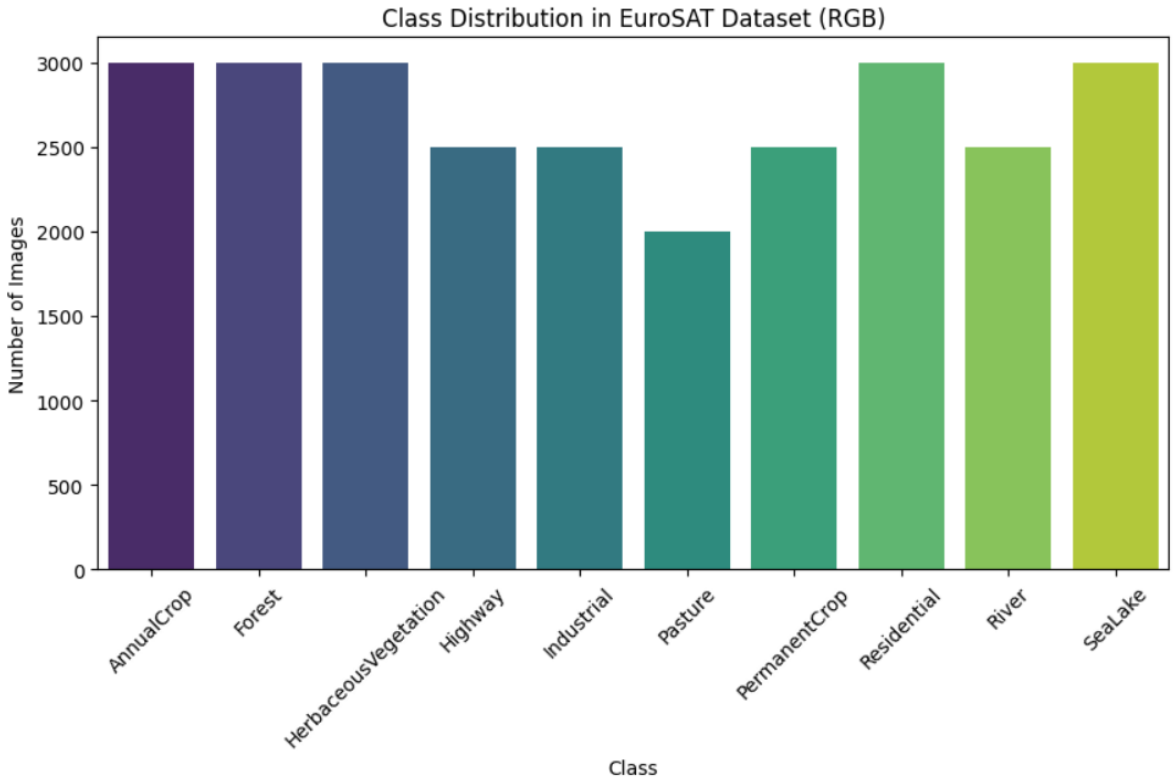

Figure 1.1: Class distribution across all 10 classes in the EuroSAT RGB dataset.

The dataset is mostly balanced across all classes, with class counts summarized in the following table:

Class	**Number of Images**
AnnualCrop	3000
Forest	3000
HerbaceousVegetation	3000
Highway	2500
Industrial	2500
Pasture	2000
PermanentCrop	2500
Residential	3000
River	2500
SeaLake	3000

Table 1.1: Image count per class.

Chapter 2

Sample Images

2.1 Random Samples from Each Class

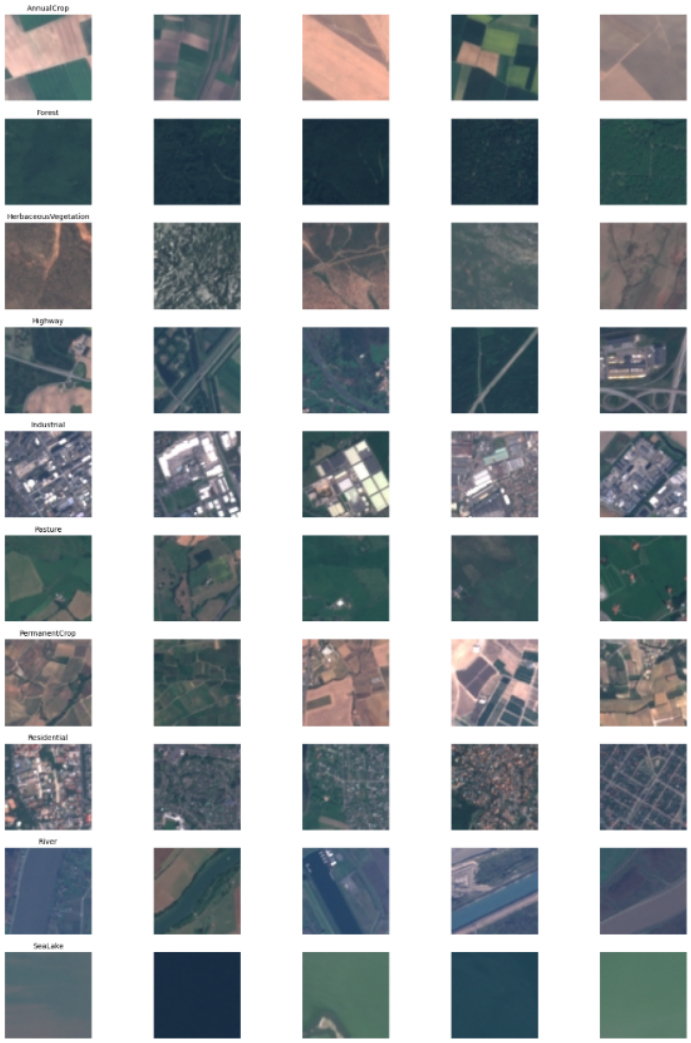

Figure 2.1: Random samples from all classes (5 samples per class).

Chapter 3

Pixel Intensity Statistics

To better understand the image distribution, channel-wise mean and standard deviation were computed across a subset of 200 images per class.

3.1 Channel Statistics

Channel	Mean Intensity	Std Intensity
Red	87.18	23.48
Green	96.73	16.82
Blue	103.53	14.27

Table 3.1: Mean and standard deviation per color channel.

3.2 Pixel Intensity Histograms

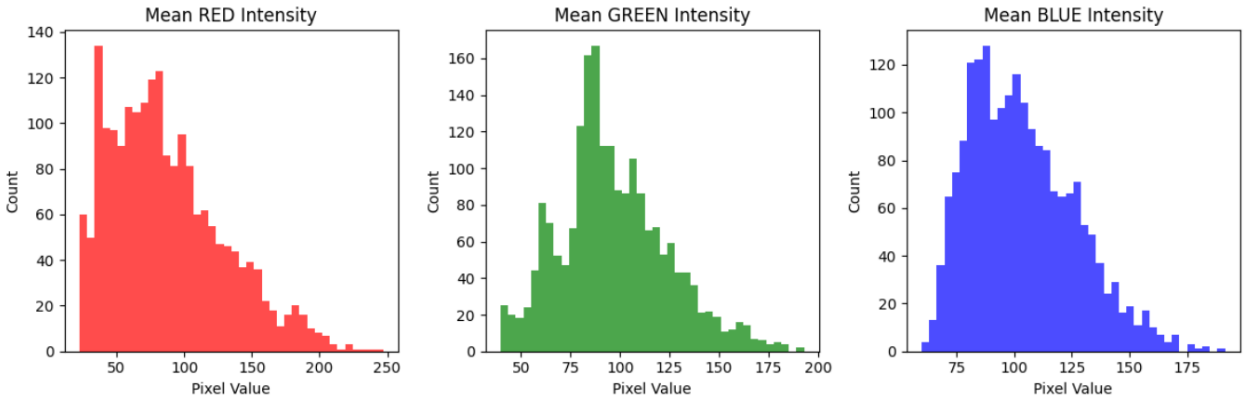

Figure 3.1: Histograms of mean pixel intensities for Red, Green, and Blue channels.

Chapter 4

Vegetation and Color Proxy Scores

Color-based proxies help characterize Earth–observation scenes. Three proxies were computed for each class:

$$\text{Vegetation Proxy} = \frac{G}{R+B}$$

$$\text{Red Proxy} = \frac{R}{G+B}$$

$$\text{Blue Proxy} = \frac{B}{R+G}$$

4.1 Vegetation Proxy (G / (R + B))

Class	Vegetation Proxy Score
Pasture	0.580
Forest	0.561
River	0.537
SeaLake	0.521
Highway	0.517
AnnualCrop	0.502
PermanentCrop	0.500
HerbaceousVegetation	0.493
Industrial	0.492
Residential	0.492

Table 4.1: Vegetation proxy score per class.

4.2 Red Proxy (R / (G + B))

Class	Red Proxy Score
AnnualCrop	0.519
PermanentCrop	0.508
HerbaceousVegetation	0.475
Industrial	0.458
Residential	0.444
Highway	0.422
River	0.344
Pasture	0.341
Forest	0.276
SeaLake	0.275

Table 4.2: Red proxy score per class.

4.3 Blue Proxy (B / (R + G))

Class	Blue Proxy Score
SeaLake	0.823
Forest	0.744
River	0.669
Pasture	0.621
Highway	0.591
Residential	0.579
Industrial	0.562
HerbaceousVegetation	0.547
PermanentCrop	0.505
AnnualCrop	0.503

Table 4.3: Blue proxy score per class.

Chapter 5

Visually Confusing Class Pairs

Some classes share similar textures and colors, making them harder for a classifier:

- AnnualCrop vs PermanentCrop
- Residential vs Industrial
- River vs SeaLake

Figure 5.1: Visualization of commonly confused class pairs.

Chapter 6

Conclusion

The preprocessing analysis highlights the dataset balance, color distribution, and spectral characteristics across land–cover classes. These insights support model training, data normalization, and understanding of class confusion patterns.