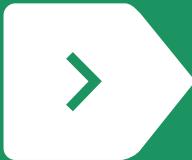


● DEPI Project

# Land Type classification & Captioning



# Agenda

---

**Milestone 1:** Data Collection, Exploration, and Preprocessing

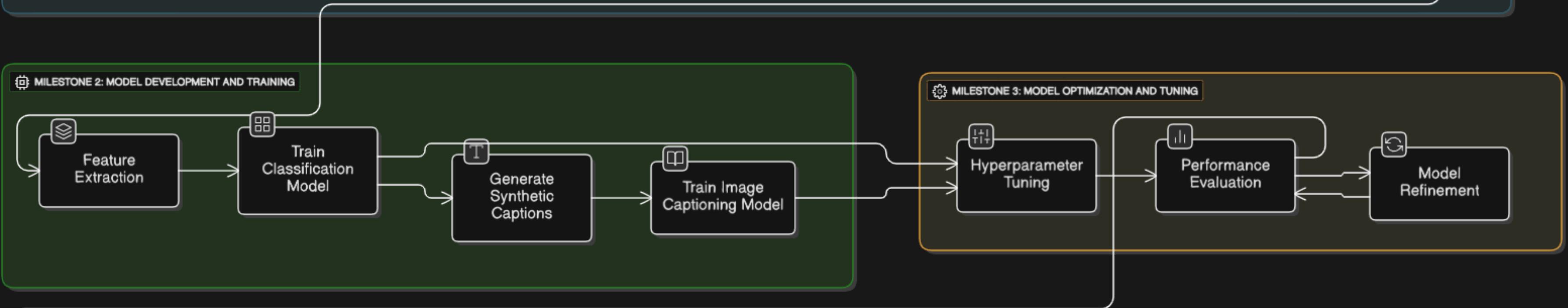
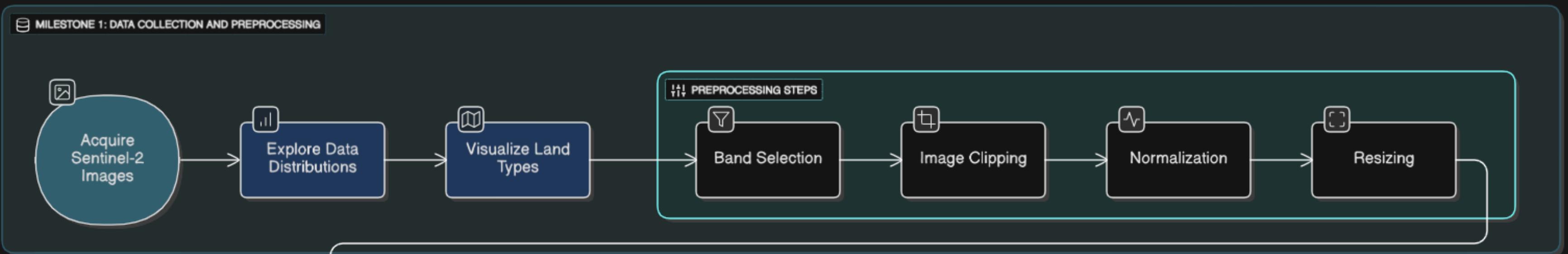
**Milestone 2:** Model Development and Training

**Milestone 3:** Model Optimization and Performance Tuning

**Milestone 4:** Real-Time Model Deployment and Visualization

**Milestone 5:** Final Documentation and Presentation





# Project Overview

---

The Land Type Classification using Sentinel-2 Satellite Images project focuses on building a **deep neural network (DNN) model** to classify major land types in Egypt from **Sentinel-2 satellite imagery**.

In addition to classification, the project also explores **image caption generation**. Instead of only predicting a single land type label, the system can generate a **descriptive natural language caption** about the scene.



# EuroSAT Dataset

---

This dataset is being used for classifying the use of land in geospatial images. The end goal for the classification is that the top 2 uses of land in an image are given as output to the user.

Kaggle DataSet Link 



# Dataset Structure

## 1] EuroSAT (RGB)

- 10 class folders
- train.csv / test.csv / validation.csv
- label\_map.json

## 2] EuroSATallBands (Multispectral)

- 10 class folders (.tif)
- train.csv / test.csv / validation.csv
- label\_map.json

There are 2 folders, namely,

- **EuroSAT** → Contains RGB images collected from the Sentinel Dataset.
- **EuroSATallBands** → Contains .tif files which have all the bands of the spectrum as collected from the Sentinel-2 satellite.

The 2 directories containing the following **class folders** :

1. Annual Crop
2. Forest
3. Herbaceous Vegetation
4. Highway
5. Industrial
6. Pasture
7. Permanent Crop
8. Residential
9. River
10. Sea Lake



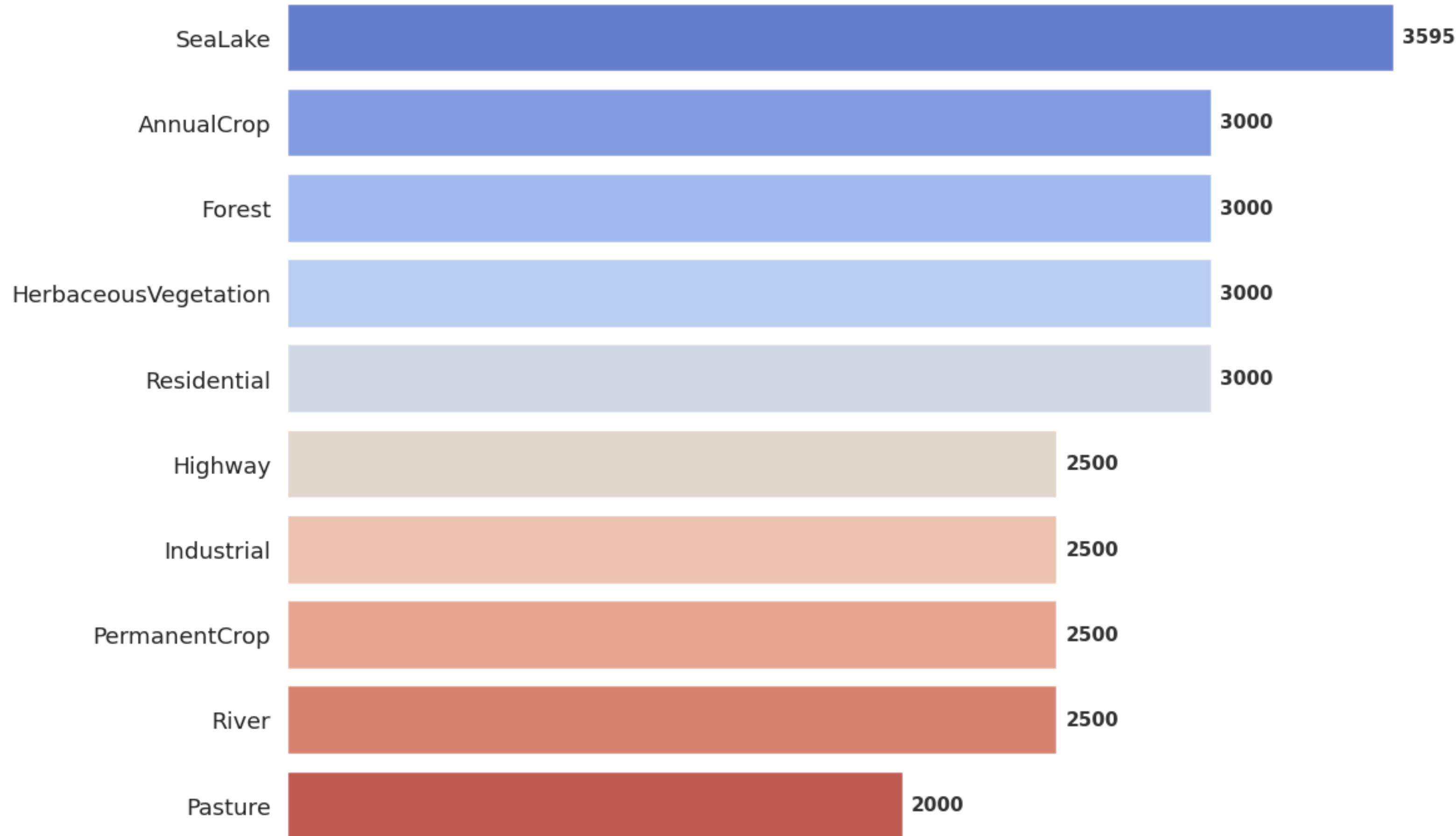
# Dataset Summary

- Total Images: 27,000
- Each image is 64x64 pixels
- Ground Sampling Distance (GSD): 10 meters per pixel
- Distribution: ~2,000 – 3,000 images per class

- AnnualCrop: 3000
- Forest: 3000
- HerbaceousVegetation: 3000
- Highway: 2500
- Industrial: 2500
- Pasture: 2000
- PermanentCrop: 2500
- Residential: 3000
- River: 2500
- SeaLake: 3000



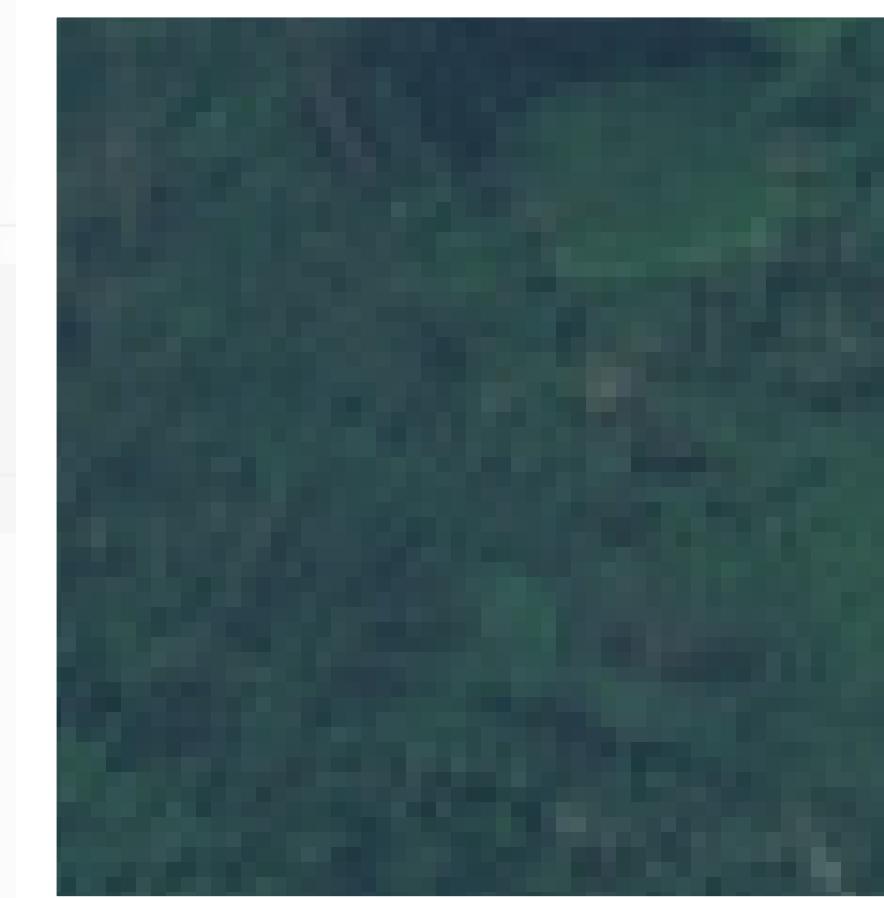
## EuroSAT Class Distribution



# Dataset Samples (RGB)



Annual  
Crop



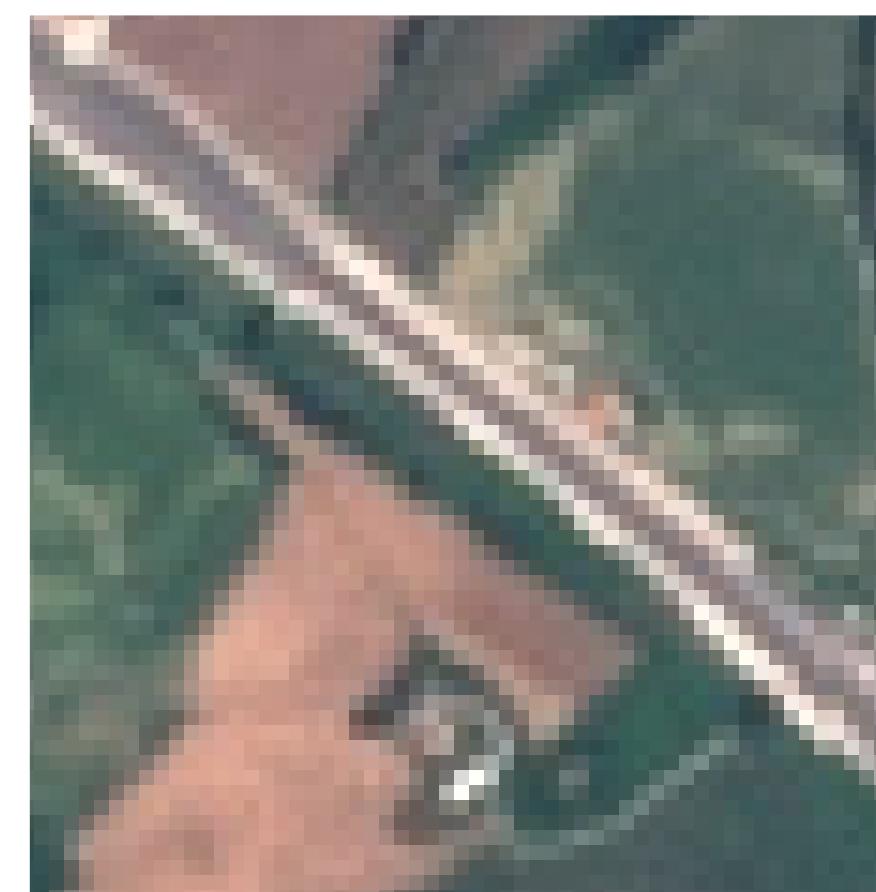
Forest



Herbaceous  
Vegetation



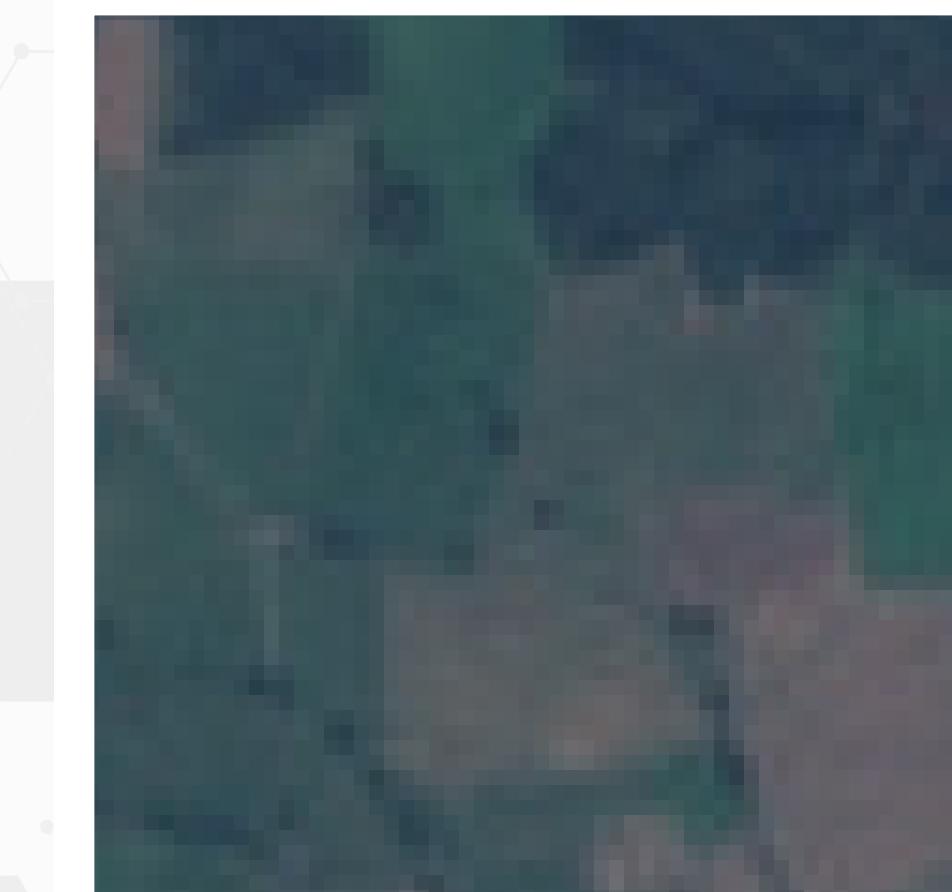
# Dataset Samples (RGB)



Highway



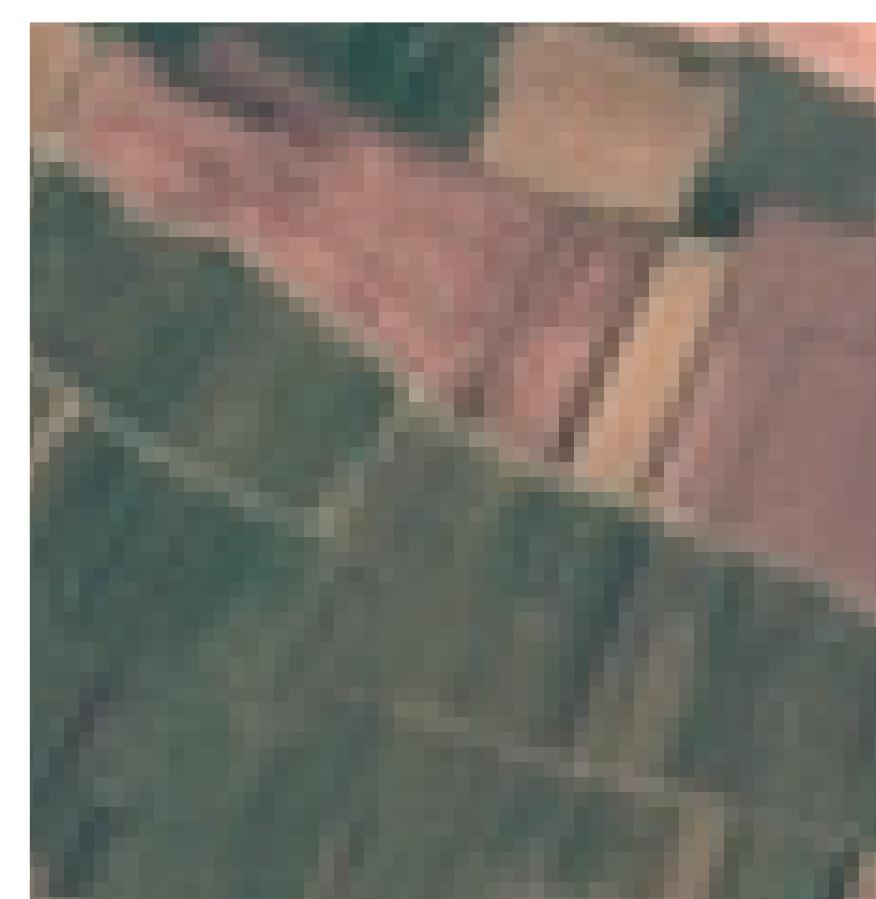
Industrial



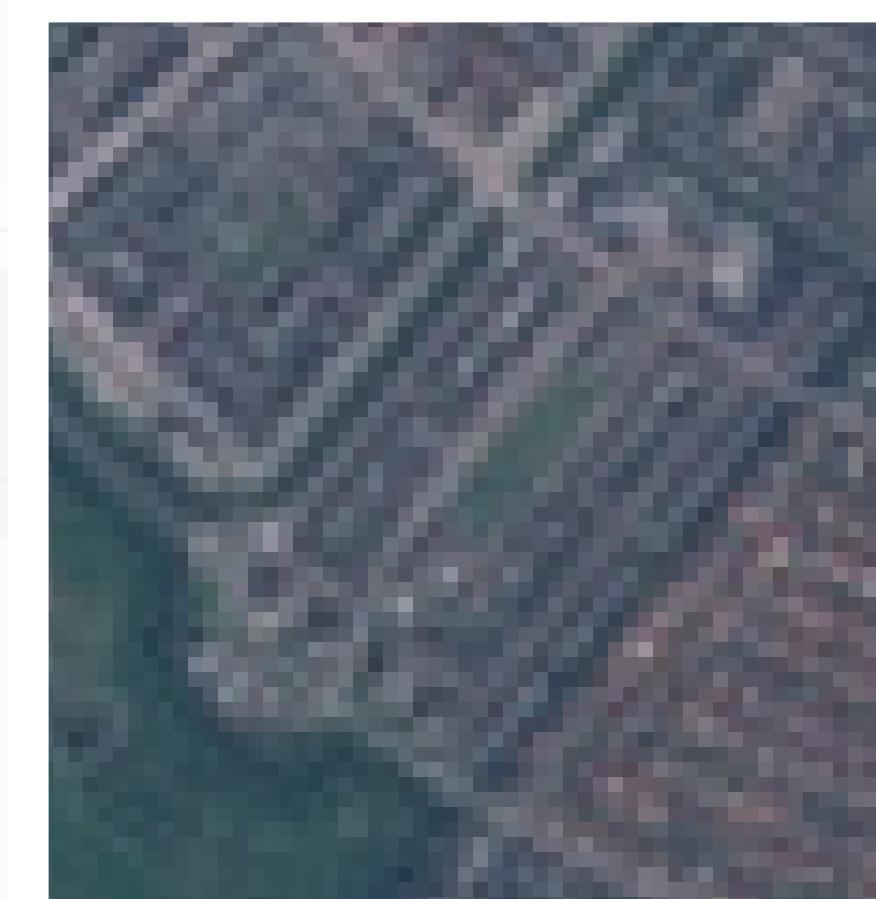
Pasture



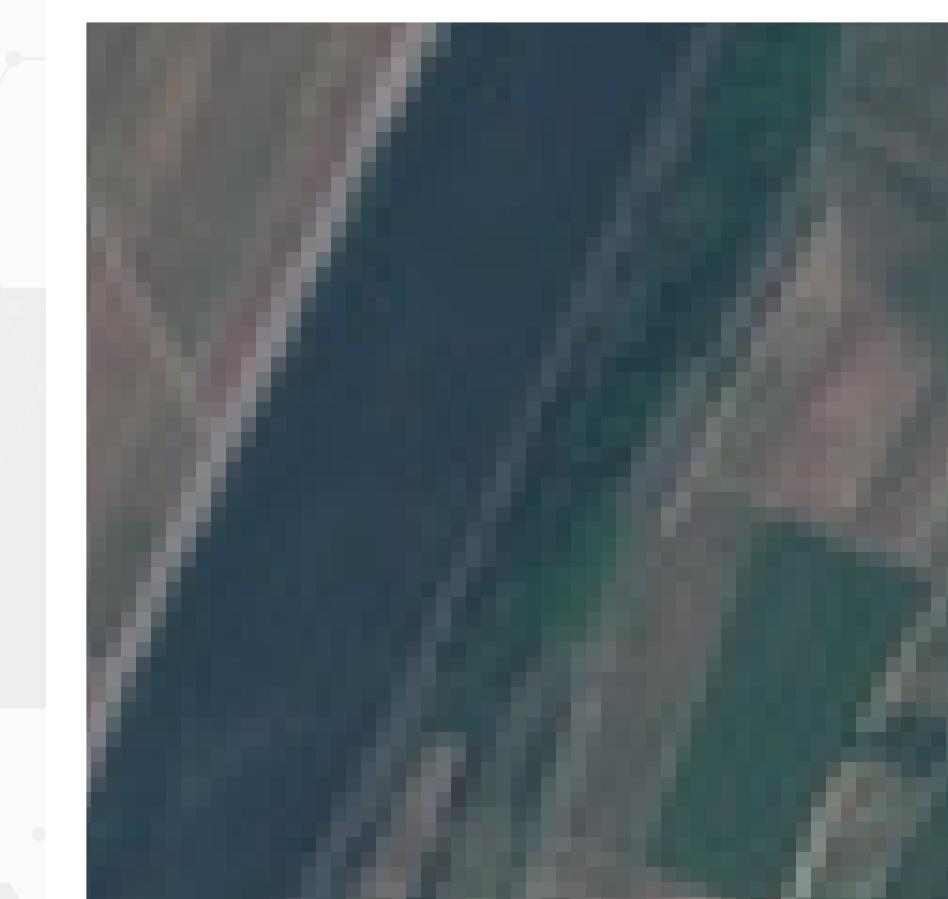
# Dataset Samples (RGB)



Permanent  
Crop



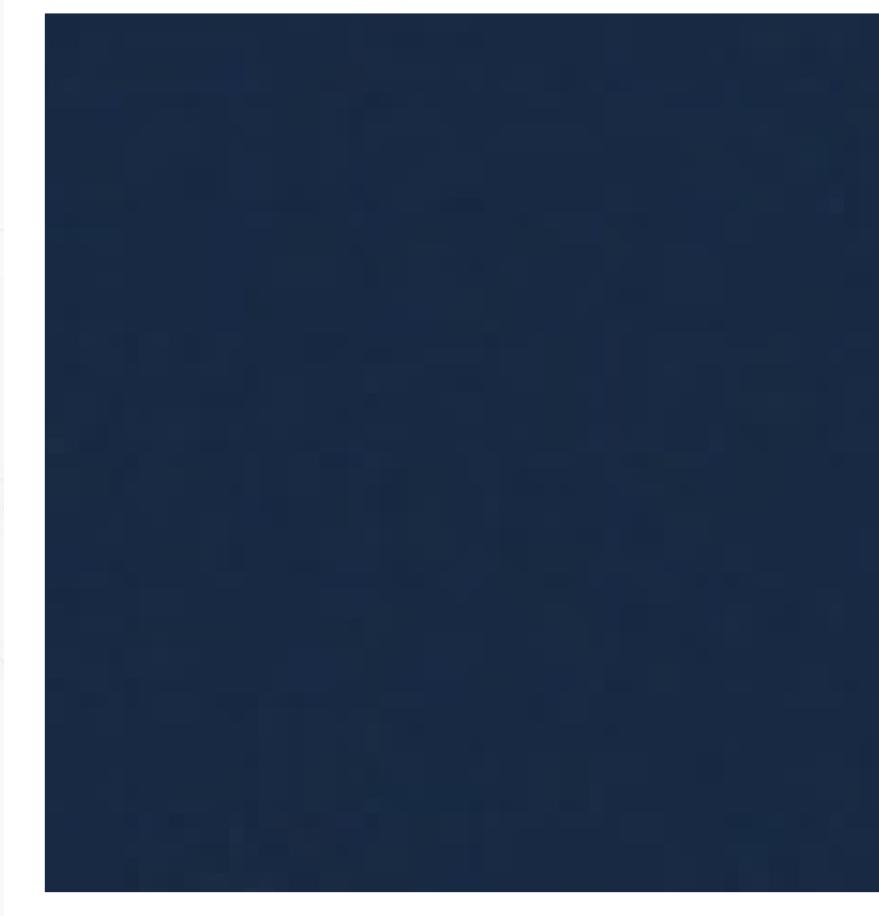
Residential



River



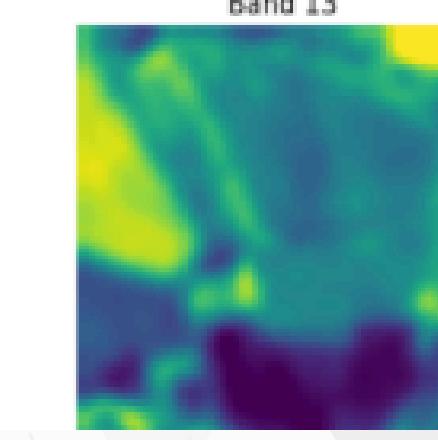
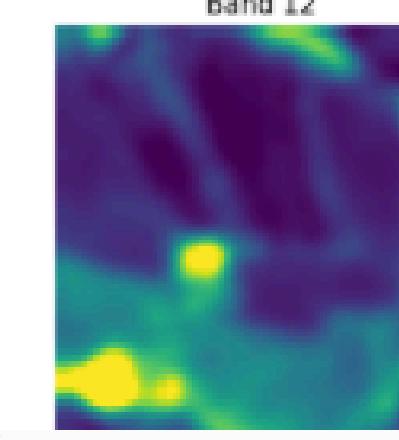
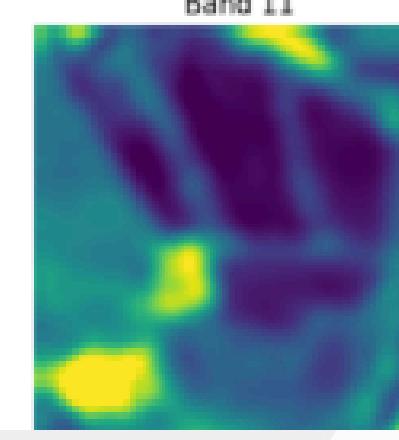
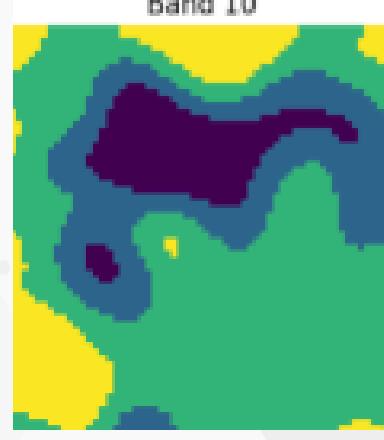
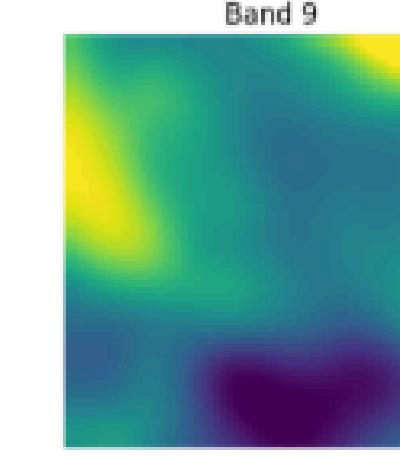
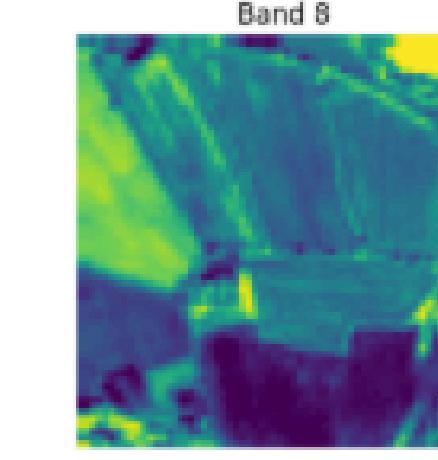
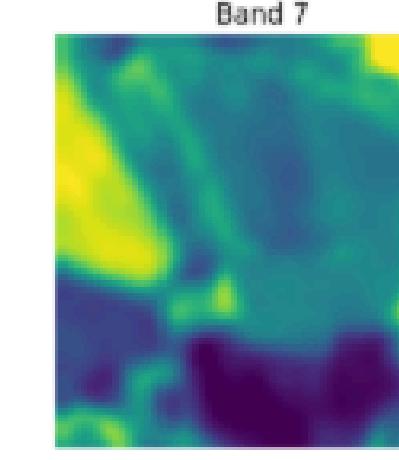
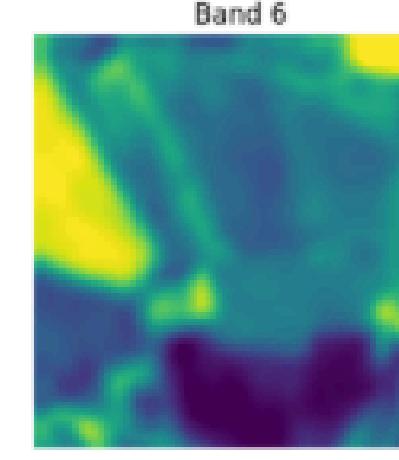
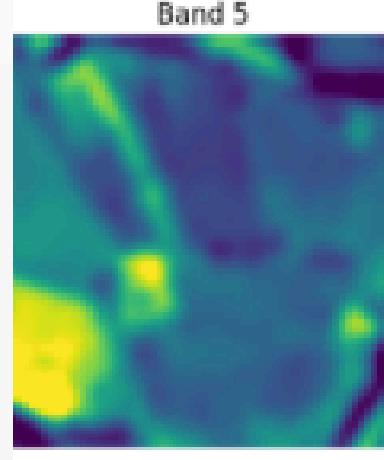
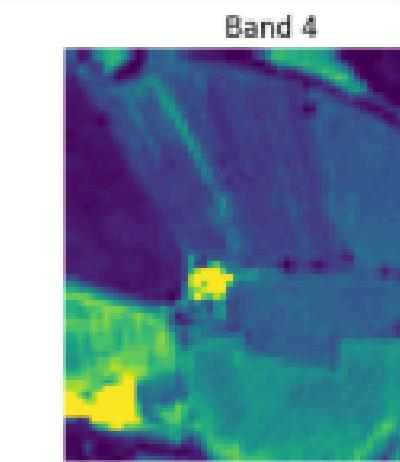
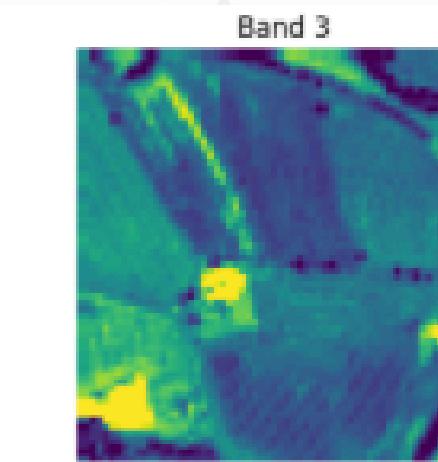
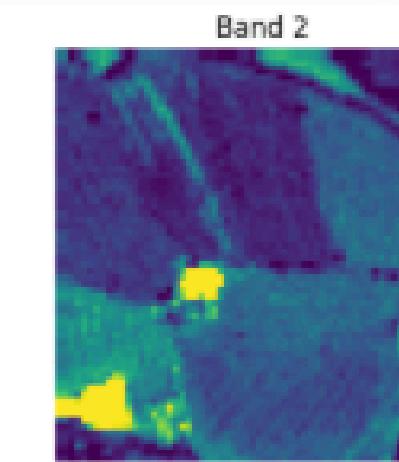
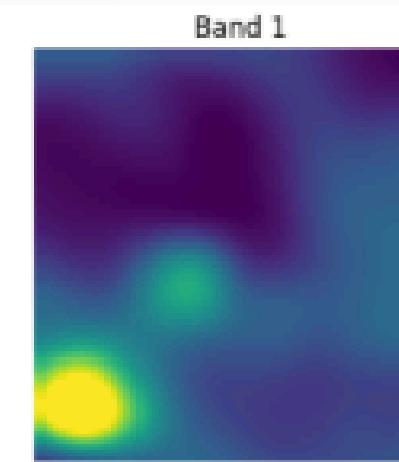
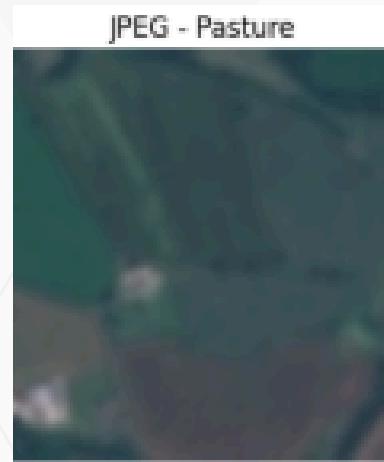
# Dataset Samples (RGB)



Sea Lake



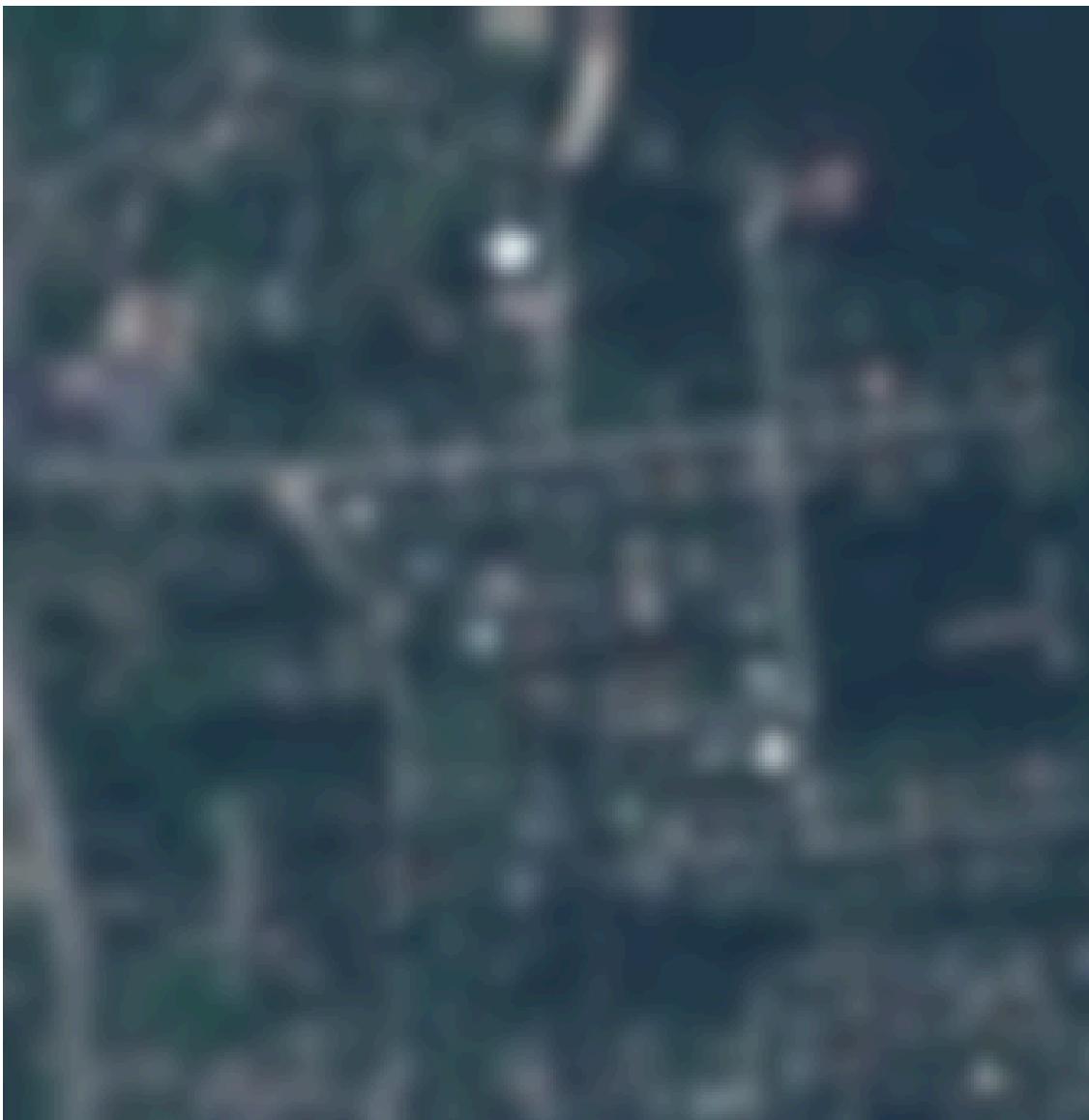
# Sample Image (13 Bands)



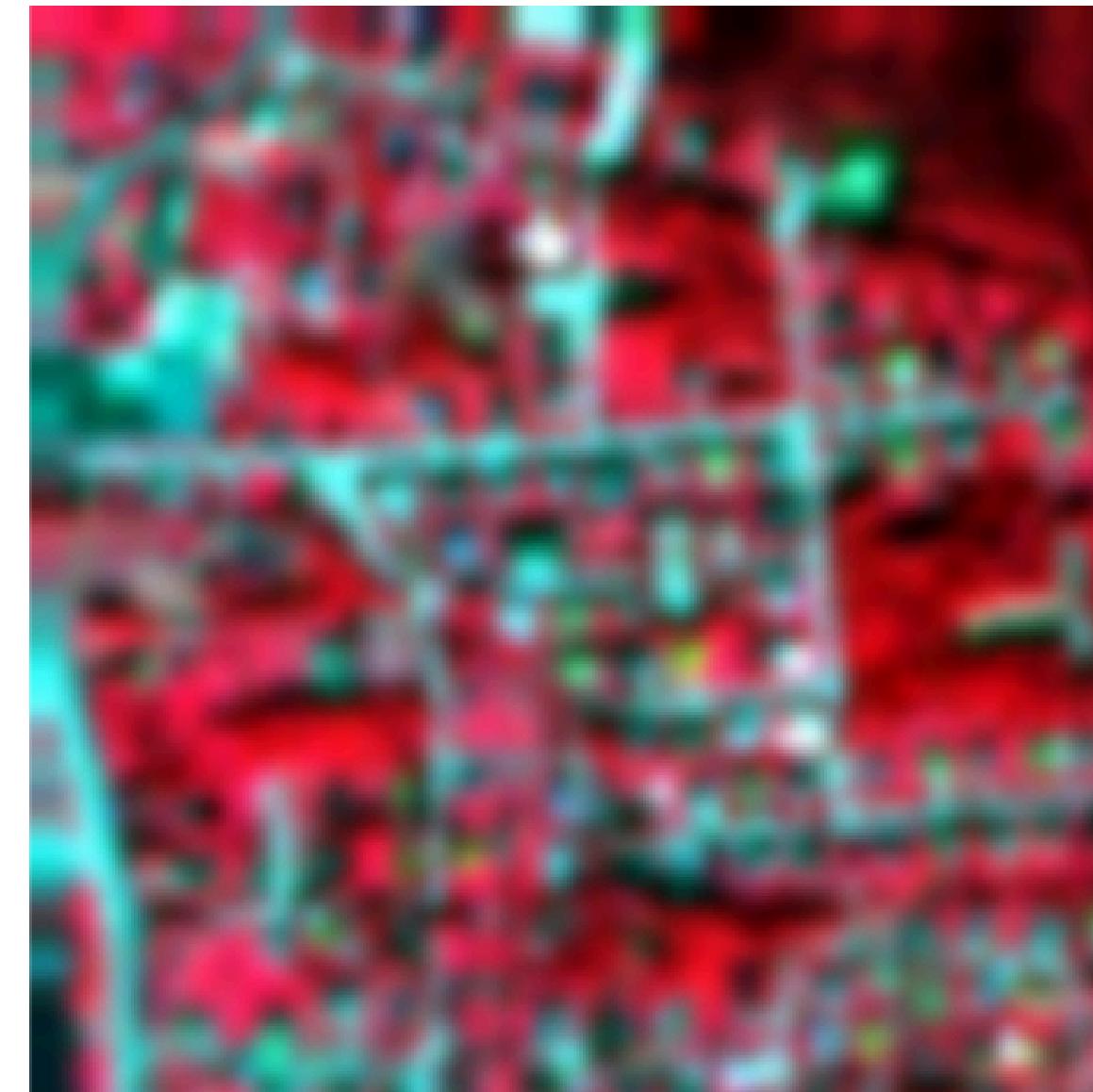
# Sample Image (NIR Red Green)

---

RGB (JPEG)



False Color (NIR-Red-Green)



# Preprocessing Pipeline

---

**Step 1:** Image Resizing

**Step 2:** Image Normalization

**Step 3:** Image Augmentation

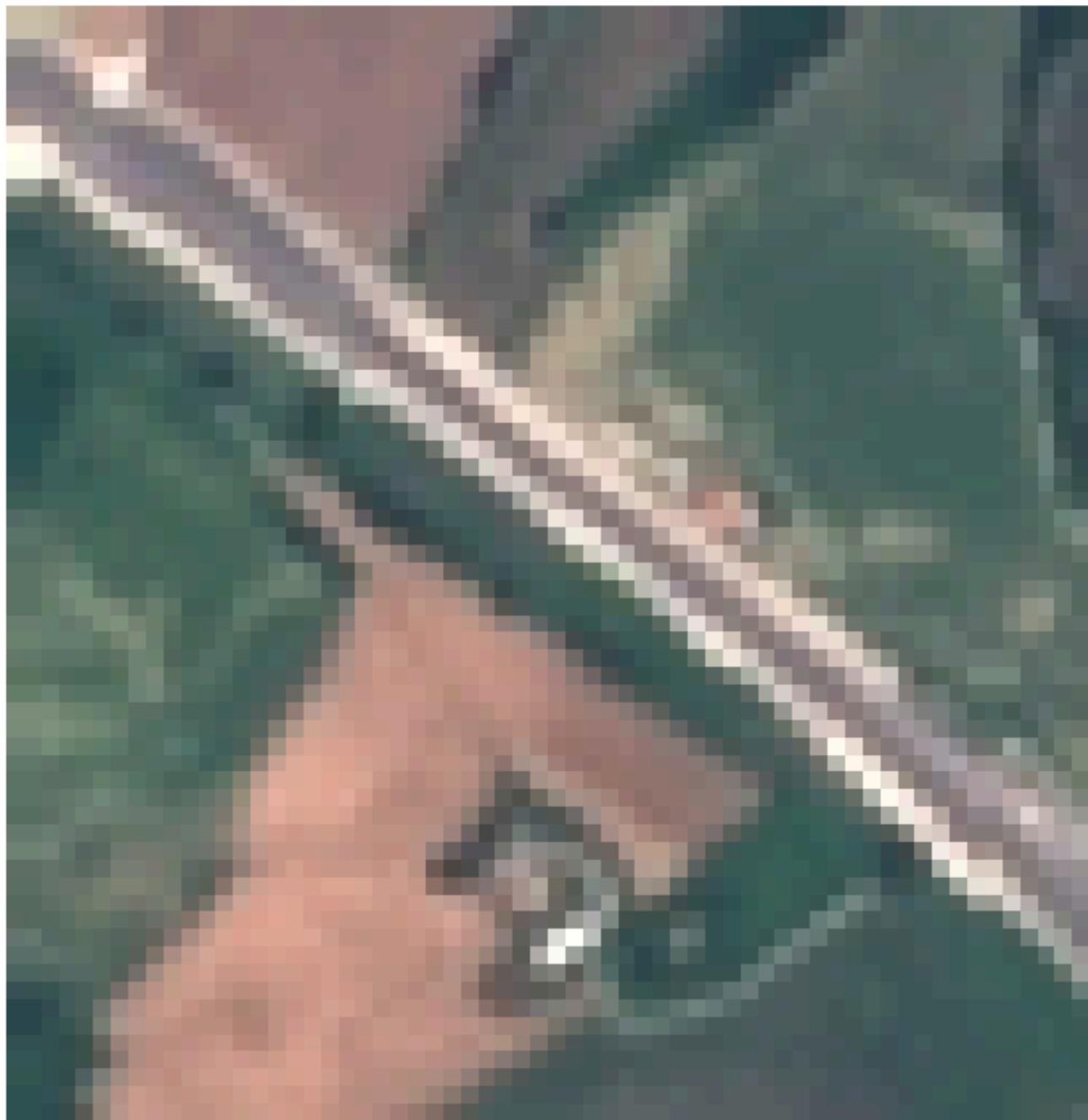
**Step 4:** Classes Encoding

**Step 5:** Dimensionality Reduction (PCA)

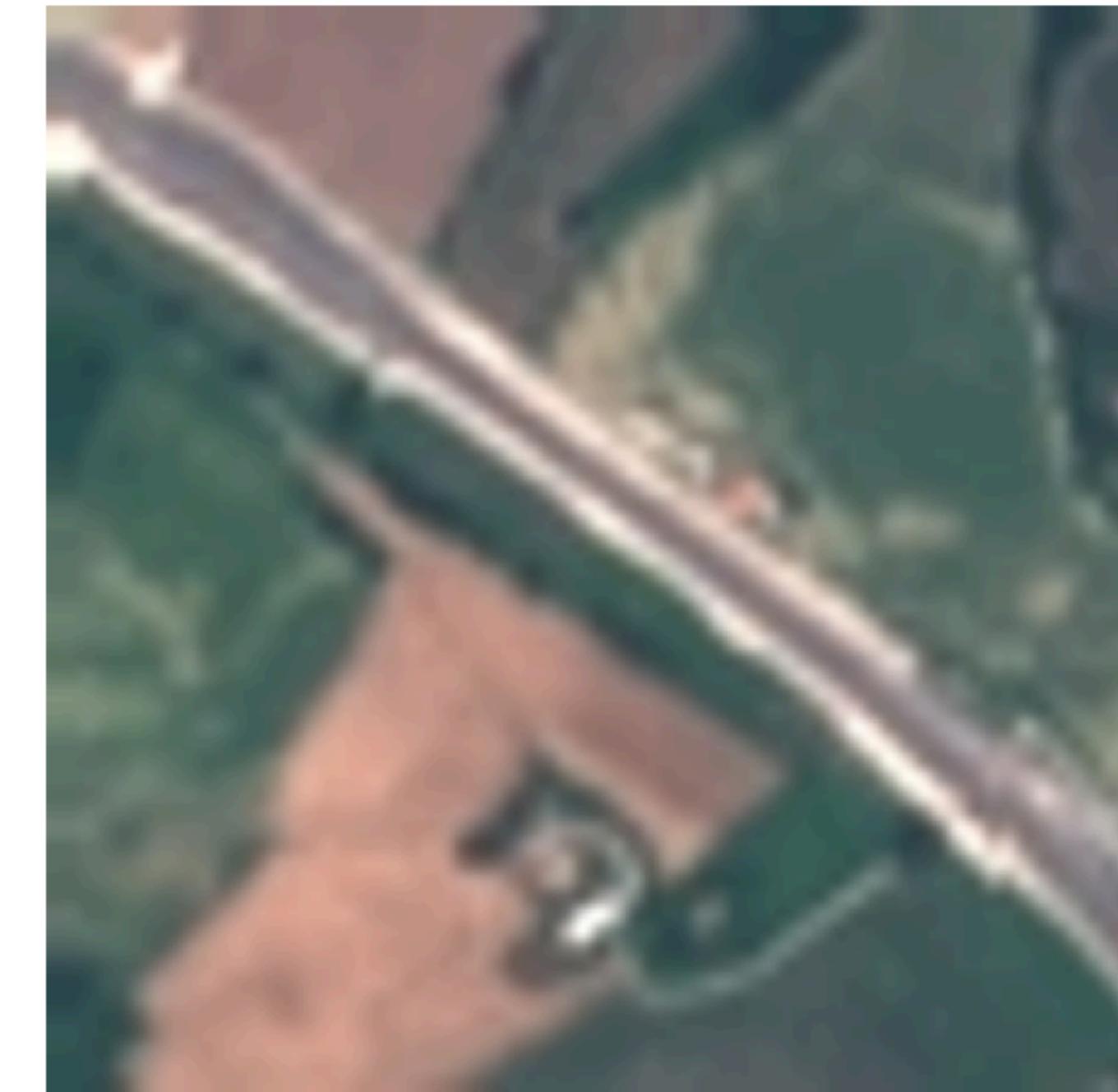


# Image Resizing

Original (64 x 64)



Resized (224 x 224)

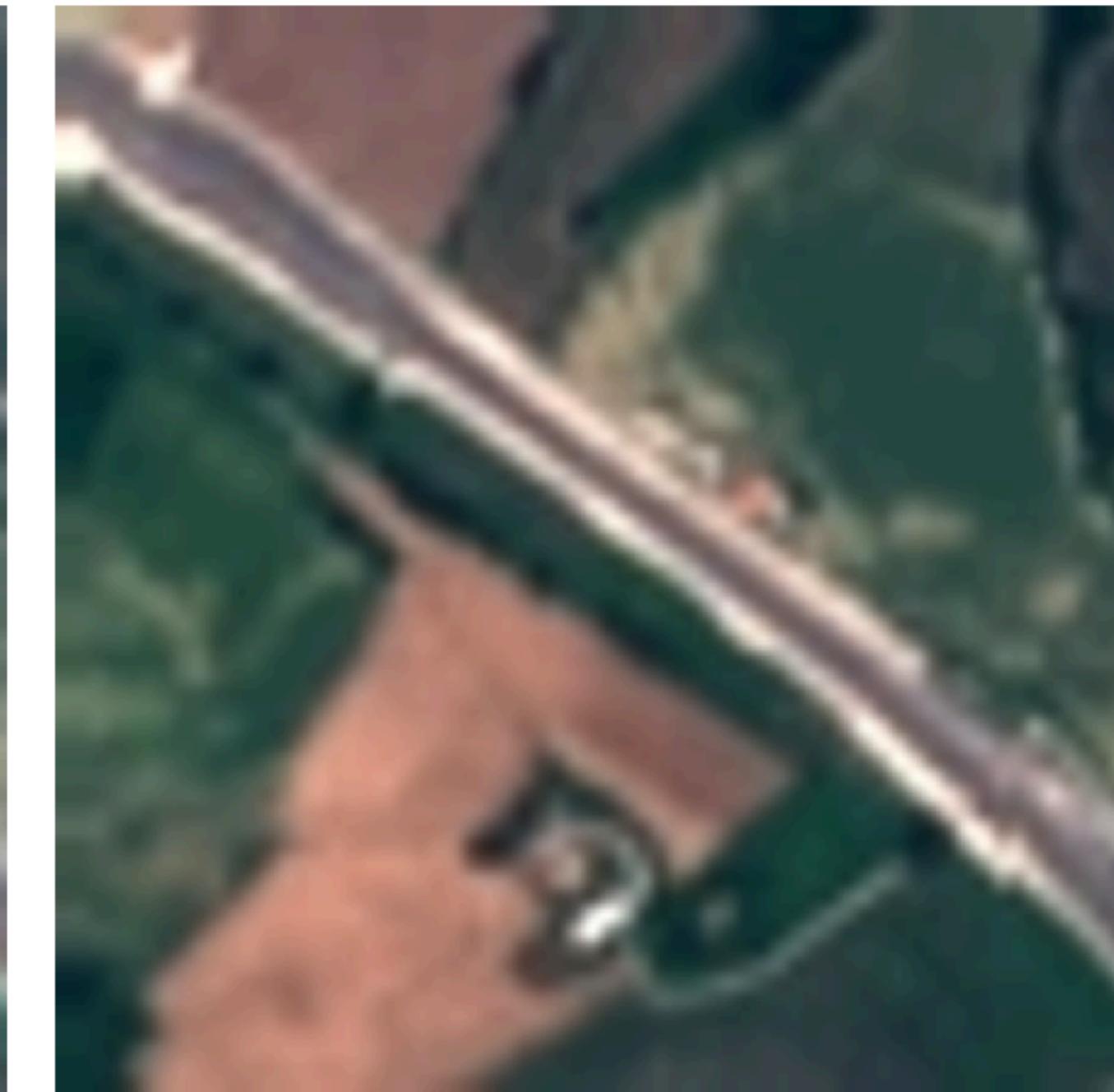


# Image Normalization

Resized (224 x 224)



Normalized (224 x 224)



# Image Augmentation

- Horizontal Flip
- Vertical Flip
- Rotation ( $0^\circ, 90^\circ, 180^\circ, 270^\circ$ )
- Brightness Adjustment
- Contrast Adjustment

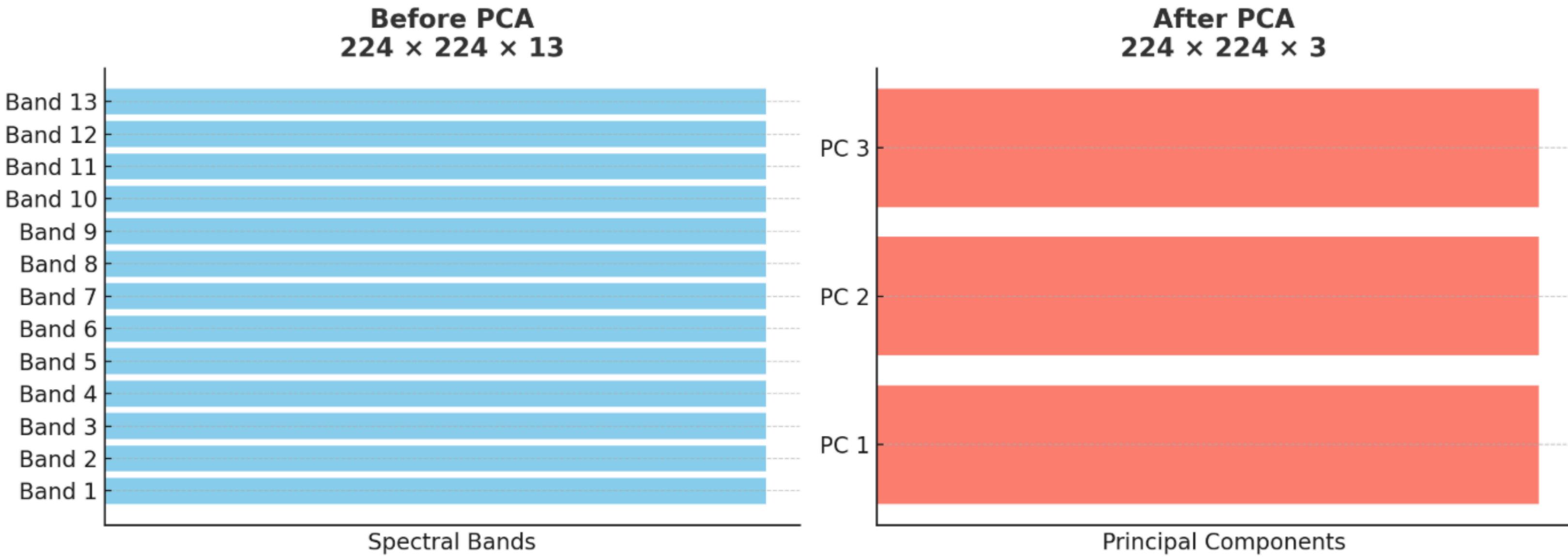


# Classes Encoding

Class Name	Label ID
Annual Crop	0
Forest	1
Herbaceous Vegetation	2
Highway	3
Industrial	4
Pasture	5
Permanent Crop	6
Residential	7
River	8
Sea Lake	9

# Dimensionality Reduction

## Dimensionality Reduction with PCA



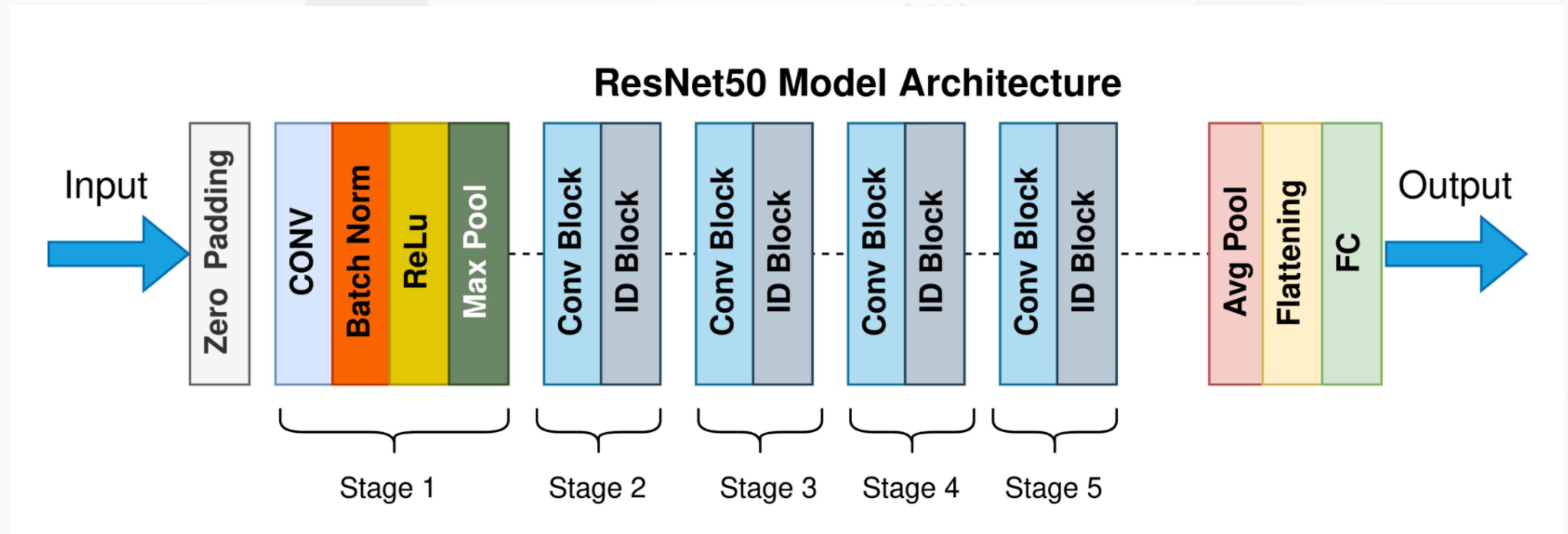
# Classification Model

- ResNet-50 has **50 layers**, and is widely used because it balances performance and efficiency.
- The model has already learned powerful feature representations from ImageNet (1.2M images, 1000 classes).

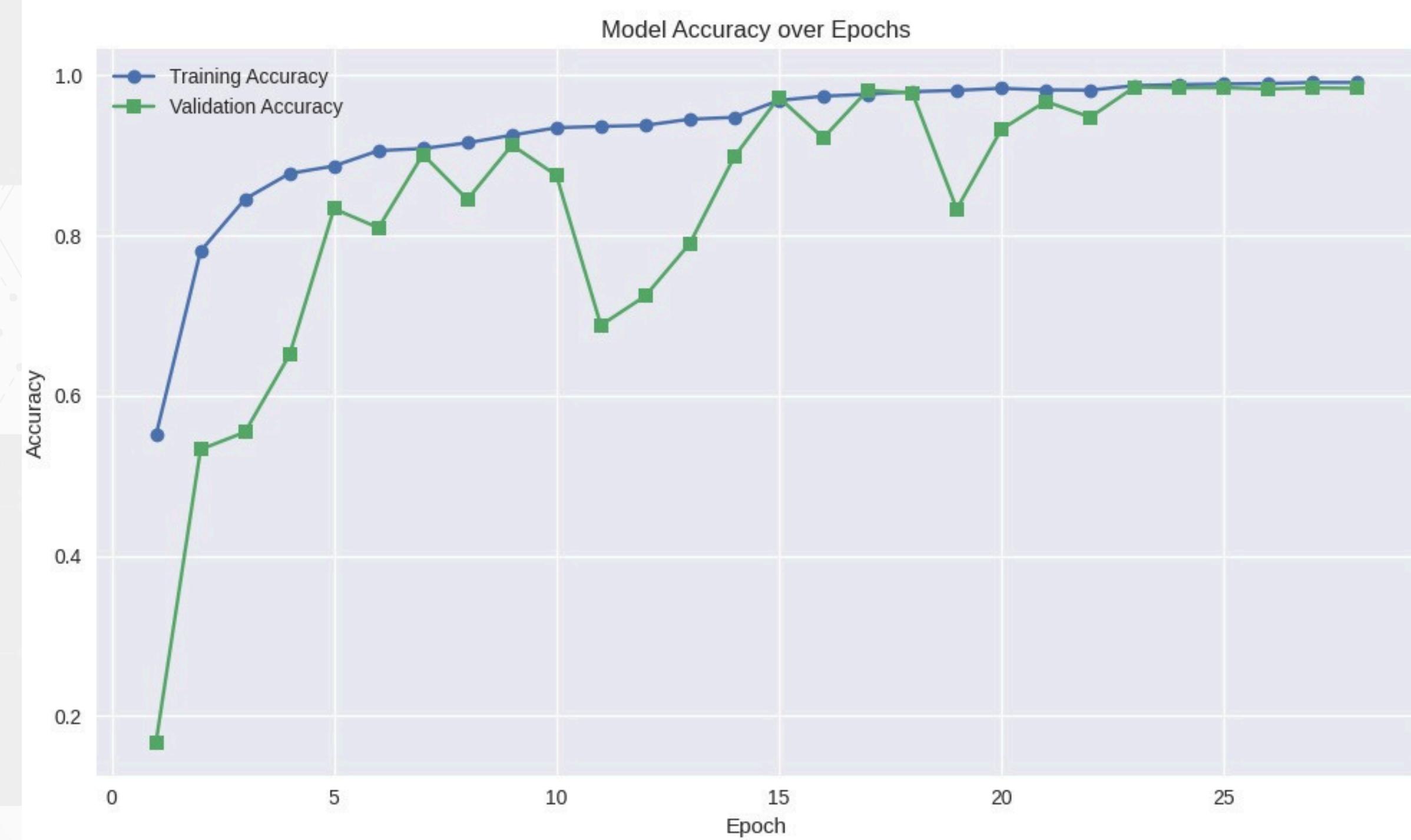
**ResNet** is a deep convolutional neural network (CNN) architecture. Its main idea is to make training very deep networks possible by solving the **vanishing gradient problem using residual connections** (or skip connections).



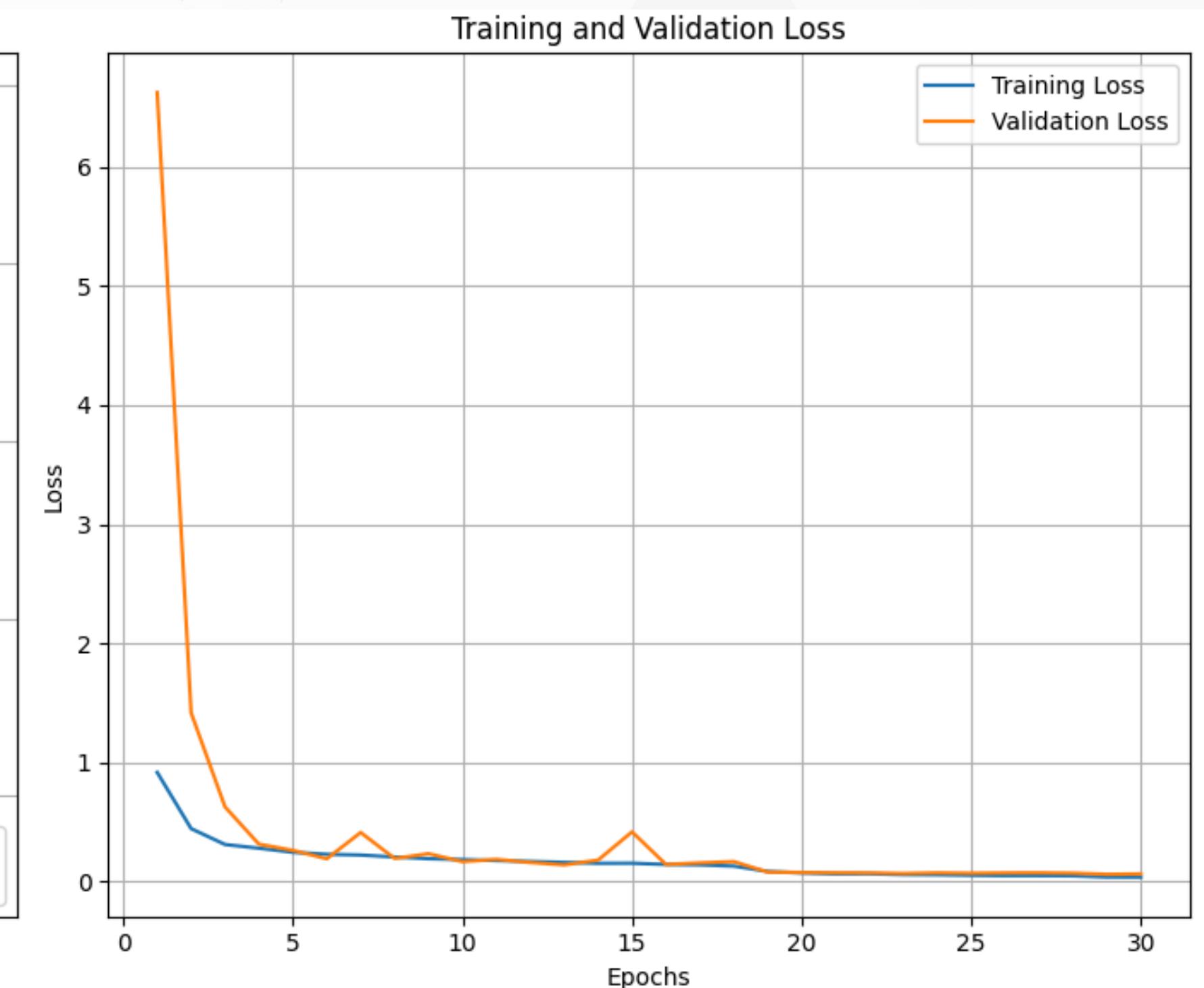
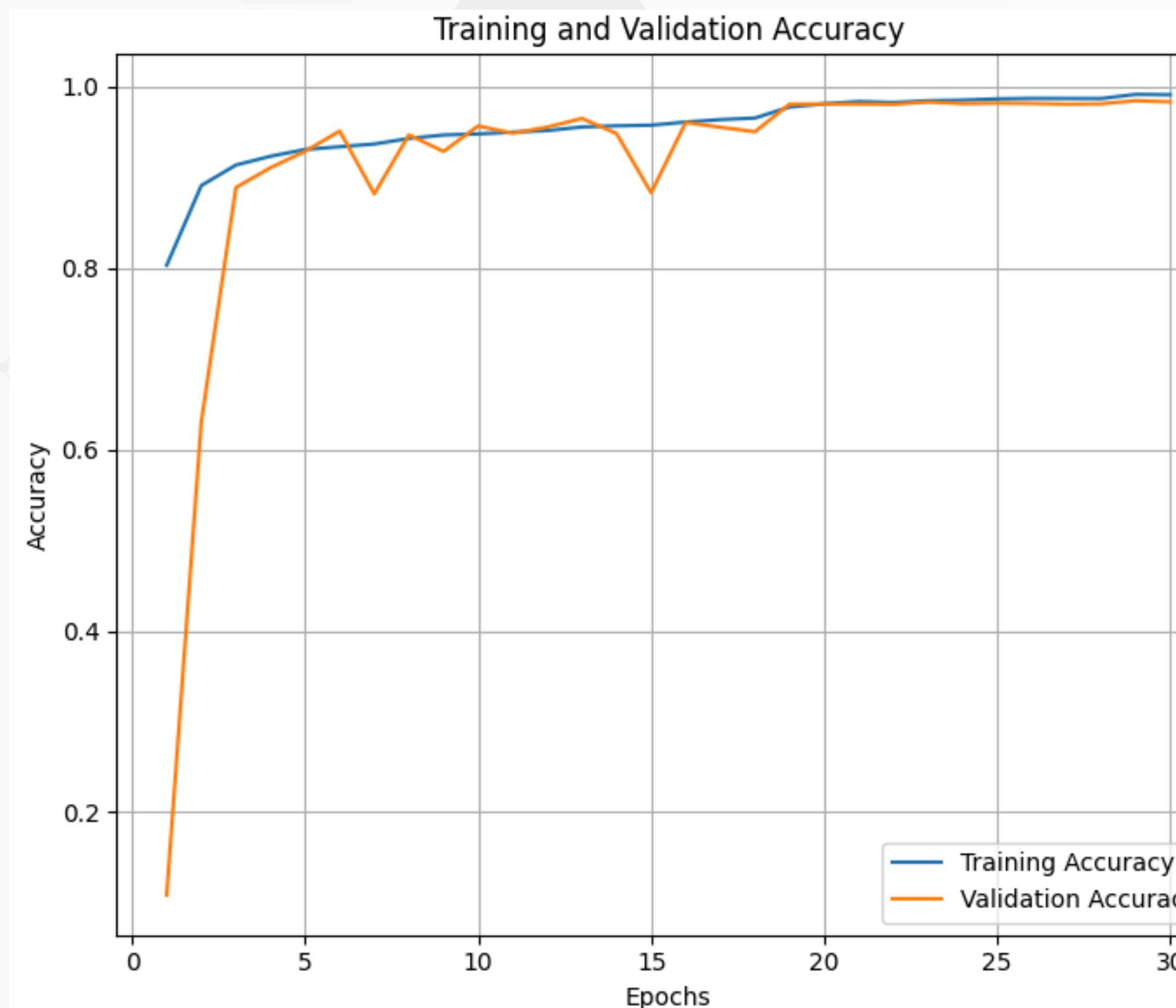
# Model Architecture



# Learning Curve (Model Adjustment)



# Learning Curve (PCA)



# Classification Report

Classification Report:

		precision	recall	f1-score	support
y_true	AnnualCrop	0.98	0.96	0.97	300
	Forest	0.99	1.00	1.00	300
	HerbaceousVegetation	0.97	0.97	0.97	300
	Highway	0.99	0.98	0.99	250
	Industrial	0.98	0.99	0.99	250
	Pasture	0.96	0.97	0.97	200
	PermanentCrop	0.96	0.94	0.95	250
	Residential	0.99	0.99	0.99	300
	River	0.98	1.00	0.99	250
	SeaLake	1.00	1.00	1.00	359
accuracy				0.98	2759
macro avg		0.98	0.98	0.98	2759
weighted avg		0.98	0.98	0.98	2759

# Confusion Matrix

The confusion matrix shows the number of correct and incorrect predictions for each class by comparing the true labels with the predicted labels.

## Confusion Matrix

# Captioning Dataset

We wanted to add an image captioning feature but due to the unavailability of the dataset, we decided to create one using LLMs.

**BLIP** stands for **Bootstrapping Language-Image Pre-training**.

BLIP is designed for **image-text tasks**, such as:

- Image Captioning
- Visual Question Answering (vQA)
- Image-Text Retrieval



# Captioing Pipeline

## 1] Classify top 3 predicted classes

- 10 class folders
- train.csv / test.csv / validation.csv
- label\_map.json

## 2] Extract Spectral bands

- NDVI (Vegetation Index):
- NDWI (Water Index):
- NDBI (Built-up Index)

Then we pass the image and the domain-enriched prompt (top predicted classes and the spectral bands) to the model to generate captions.



# Captioning Example



A patch of annual crop, permanent crop, pasture with bare or very sparse vegetation, significant water presence ( rivers / lakes ), and highly urbanized or dense settlements.

# Thank You