GTC Land Type Classification

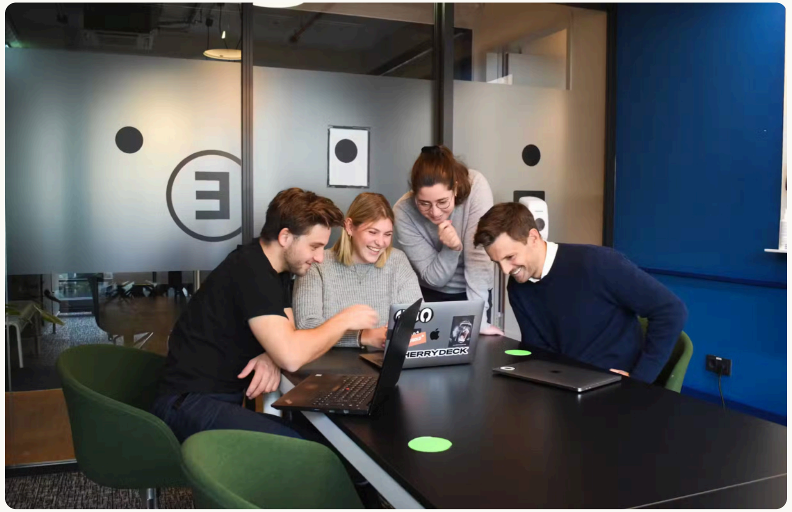
Deep Learning System for Accurate Land Classification Using Sentinel-2 Satellite Imagery

Meet the Team

GTC Team

Our dedicated team of machine learning engineers and data scientists brings together diverse expertise in computer vision, satellite imagery analysis, and web development.

Working collaboratively as part of the GTC internship program, each team member contributes specialized skills to deliver a comprehensive land classification solution.



Ahmed Maher	**Roaa Ahmed**
Hamza Abdelsalam	**Omnia Mohamed**
Alaa Ramadan	**Rowan Mohamed**

The Challenge

Critical Need

Accurate land type classification is essential for modern environmental management and urban planning. Traditional manual classification methods are time-consuming, expensive, and often inconsistent across large geographic areas.

Satellite imagery provides unprecedented coverage but presents unique challenges with high-dimensional data that requires sophisticated preprocessing and machine learning approaches.

Project Workflow Overview

EuroSAT Dataset

EuroSAT dataset description, class distribution overview, and preparation for use in model training

Data Preparation

EuroSAT dataset extraction and organization into consistent folder structure with integrity verification

EDA & Feature Building

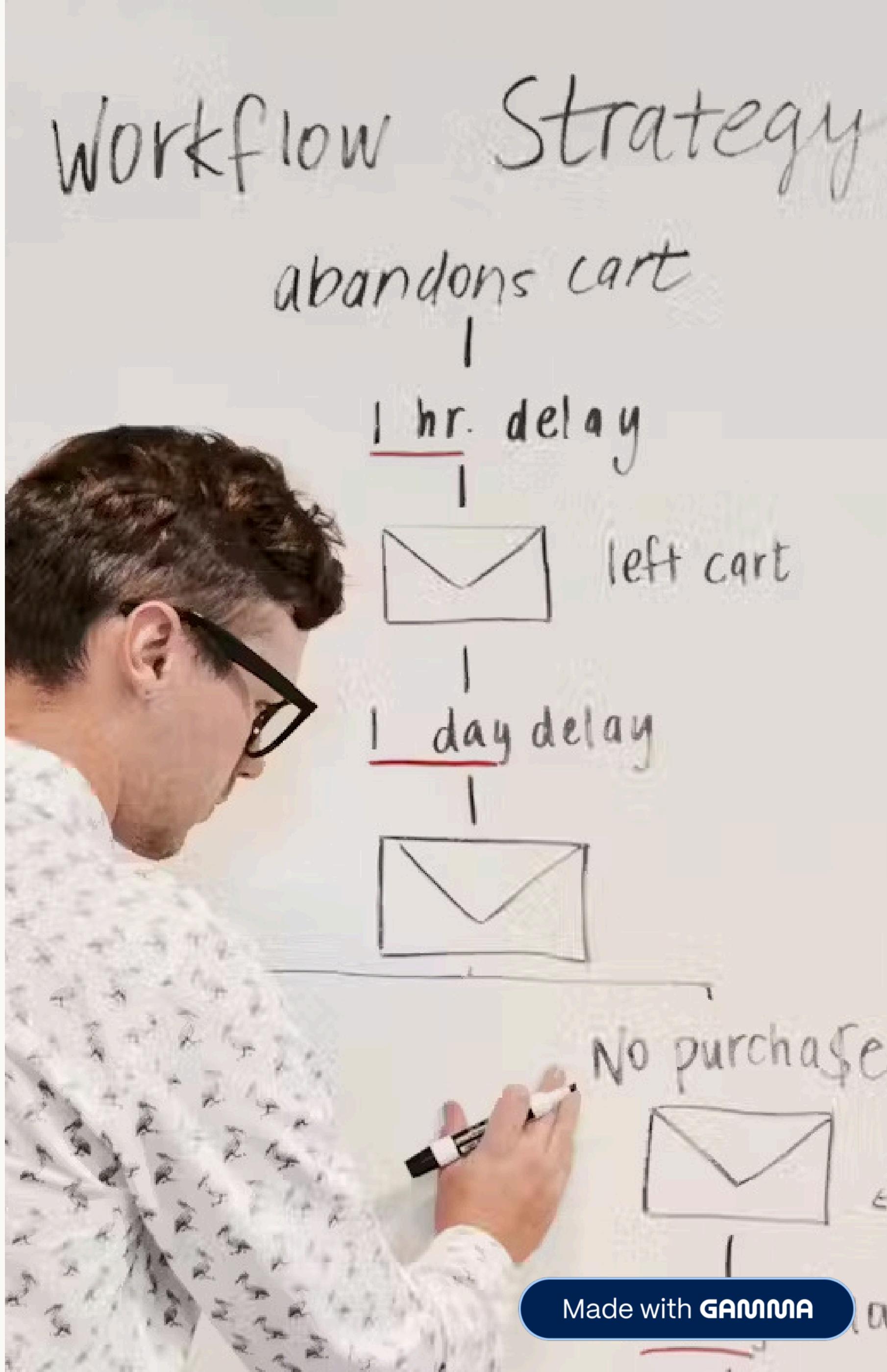
Image preprocessing, augmentation techniques, and comprehensive dataset analysis

Model Training

CNN and transfer learning model development with performance evaluation

Deployment

Web interface creation for real-time classification predictions



EuroSAT Dataset

The EuroSAT dataset is a public remote sensing dataset built from Sentinel-2 satellite images. It features 10 distinct land use/land cover classes across diverse European regions, making it an excellent resource for classification, segmentation, and change detection tasks in environmental monitoring and urban planning.

Explore the EuroSAT Dataset on Kaggle

Dataset Details

Comprising approximately 27,000 images, each 64x64 pixels, the dataset captures rich spectral information across 13 Sentinel-2 multi-spectral bands with varying resolutions.

Number of classes	10 land cover / land use categories
Number of images	~27,000 images
Image size	64 × 64 pixels
Spectral bands	13 (Sentinel-2 multispectral)
Resolution	10–60 m per pixel (band-dependent)

Land Cover Classes

- Annual Crop
- Forest
- Herbaceous Vegetation
- Highway
- Industrial
- Pasture
- Permanent Crop
- Residential
- River
- Sea / Lake

Phase 1: Data Preparation

This initial phase focused on meticulously preparing the EuroSAT dataset to ensure high-quality input for our deep learning models. It involved structured organization, image preprocessing, and strategic data augmentation to optimize model training.

Dataset Collection & Verification

Successfully collected the EuroSAT dataset, based on Sentinel-2 satellite imagery, and conducted thorough verification to ensure data integrity and accurate class labeling.

Structured Organization & Preprocessing

Organized the dataset into a consistent folder structure suitable for machine learning pipelines. Images were preprocessed and resized to a uniform 64×64 pixel dimension.

Robust Data Augmentation

Applied various data augmentation techniques, including horizontal/vertical flips, brightness adjustments, and contrast modifications, to enhance model generalization and prevent overfitting.

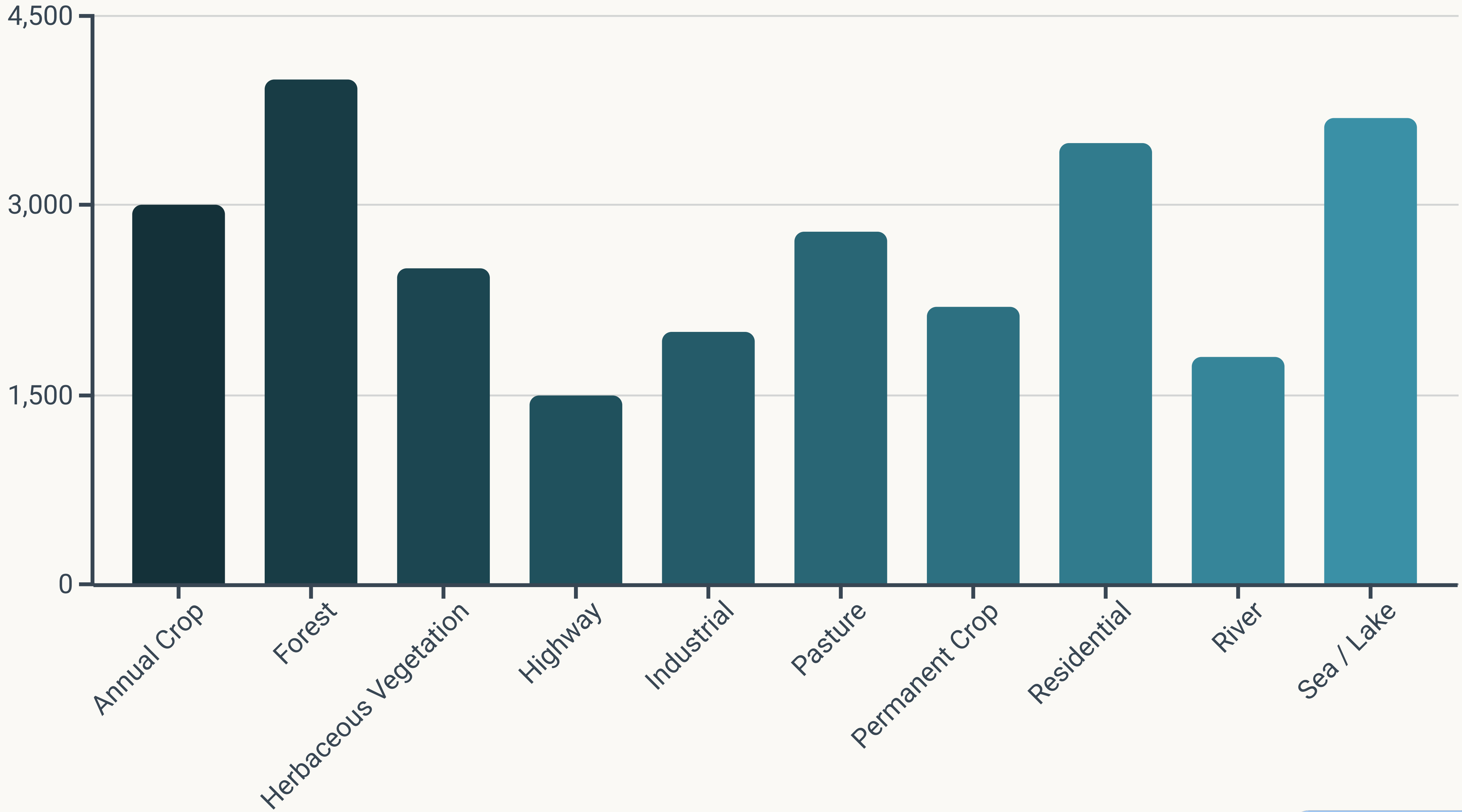
Dataset Splitting

Divided the processed dataset into an 80% training set and a 20% validation set, providing distinct data for model development and unbiased performance evaluatio

Phase 1: Data Preparation

Class Distribution

Understanding the distribution of land cover classes is crucial for balanced model training. This chart illustrates the number of images per class in the EuroSAT dataset.



Phase 2: EDA & Feature Building

This phase focused on refining our dataset through in-depth analysis, strategic preprocessing, and augmentation to prepare it for robust model training.

Class Distribution Exploration

We thoroughly explored the dataset's class distribution to identify potential imbalances and inform sampling strategies, ensuring comprehensive coverage across all land types.

Pixel Value Normalization

Pixel values were normalized across all images, standardizing input ranges to enhance model stability and accelerate convergence during training.

Robust Data Augmentation

To prevent overfitting and improve generalization, we applied various augmentation techniques including rotations, zooms, and horizontal flips, expanding the effective dataset size.

Class Imbalance Visualization

Visualized class imbalances using detailed charts, providing critical insights for implementing appropriate re-sampling or weighting techniques in subsequent model training phases.

Standardized Feature Building

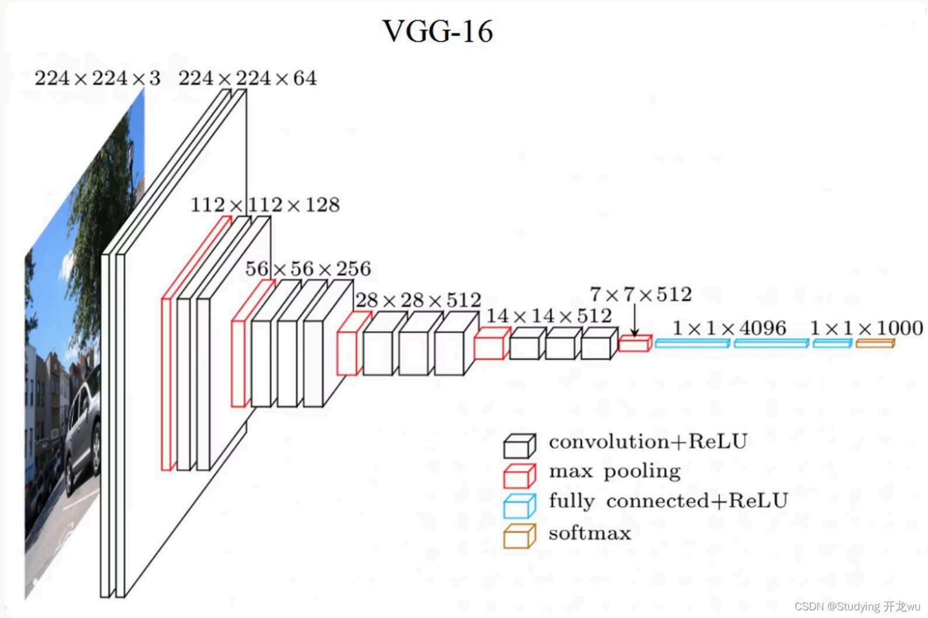
Features for model training were constructed using a standardized preprocessing pipeline, ensuring consistency and optimal input quality for deep learning architectures.

Phase 3: Model Training

This phase focused on developing and refining our classification models, leveraging both custom CNN architectures and transfer learning to achieve optimal land type classification.

Model Development & Tuning

- Built a baseline **Convolutional Neural Network (CNN)** model from scratch for initial classification.
- Applied transfer learning with **Baseline** & **VGG16**, fine-tuning the pre-trained model for improved accuracy and generalization.
- Systematically **tuned hyperparameters** (learning rate, batch size, epochs) to optimize model performance.
- Evaluated models using key metrics: **Accuracy, Precision, Recall, and F1-score**, ensuring robust performance assessment.
- Achieved strong performance on the validation set, demonstrating the model's effectiveness in land type classification.



Baseline vs VGG16 Comparison

Our comprehensive evaluation highlights the superior performance of the VGG16 model, fine-tuned through transfer learning, compared to our custom-built baseline CNN. This comparison was critical in selecting the optimal model for our land type classification task.

Overall Accuracy	88%	93%
Strongest Class	SeaLake (97%)	Residential (99%)
Weakest Class	Highway (74%)	River (86%)
Generalization	Moderate	Strong

The VGG16 model demonstrated significant improvements across all key metrics, including a 5% increase in overall accuracy and notably better performance on both its strongest and weakest classified classes. Its robust generalization capabilities make it the preferred choice for deployment.

Phase 4: Deployment

Web interface creation for real-time classification predictions, bringing our model to life for practical application.

1

Flask API Integration

The trained model was deployed as a robust Flask API, ensuring efficient and scalable real-time prediction services.

2

Intuitive Web Interface

A simple, user-friendly web interface was designed for easy satellite image upload and interaction.

3

Real-time Inference

Seamless backend integration enables instant classification of uploaded images using the deployed model.

4

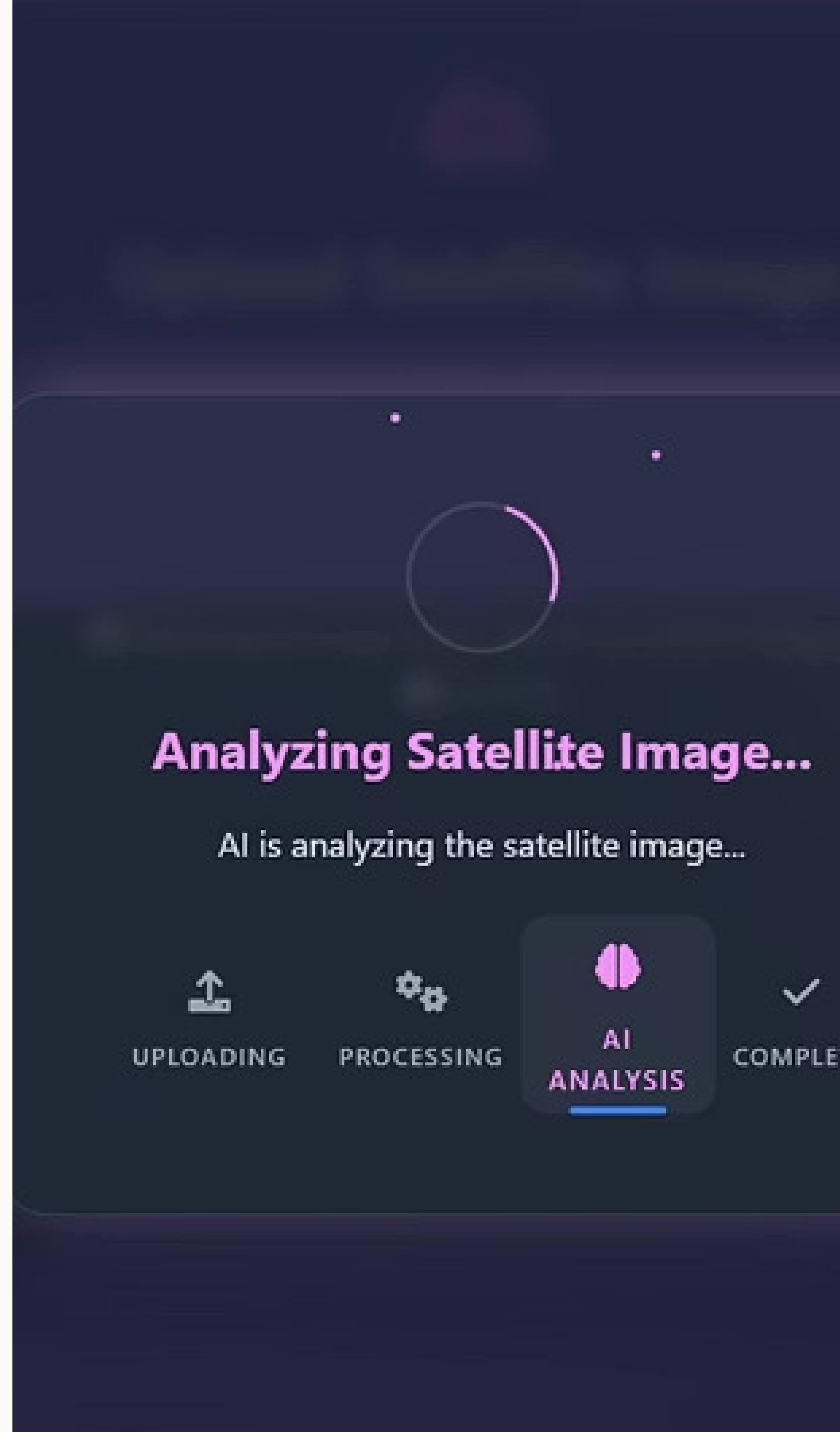
Clear Prediction Display

Predicted land types are clearly displayed along with confidence scores for transparency and user understanding.

5

Reliable User Experience

Focus on smooth interactions ensures a reliable platform for accurate and quick land type predictions.



thank you 🤍🤍