📝 *Sentinel-2 Land Classification Project Report.*

1. Project Introduction:

This project aims to utilize **Deep Neural Networks (DNNs)** to analyze and classify various land cover types by leveraging free, **multispectral satellite images** captured by the **Sentinel-2** satellite from the European Space Agency. The target land classification categories include Agricultural Lands, Water Bodies, Urban Areas, Desert, Roads, and Trees/Forests. The project serves as a vital tool to support wide-ranging applications such as smart urban planning, environmental monitoring, and resource management.
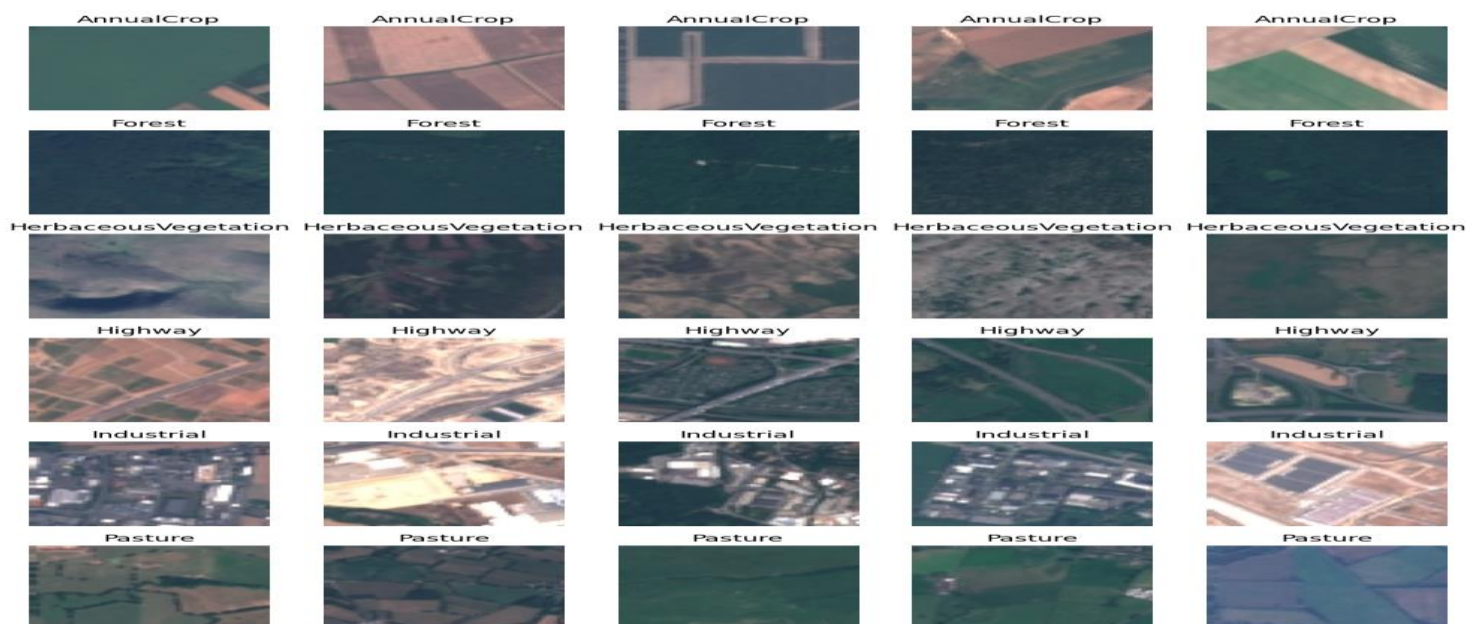
2. Data Acquisition and Preprocessing

Data Collection and Exploration

Data was gathered from official platforms like the Copernicus Open Access Hub and USGS Earth Explorer, alongside exploring ready-made options like the EuroSat Dataset. The requirement was multispectral images, including the Visible Spectrum (RGB) and the Near-Infrared (NIR) bands. Exploration involved checking for data balance, missing values, duplicates, and mislabeled images.
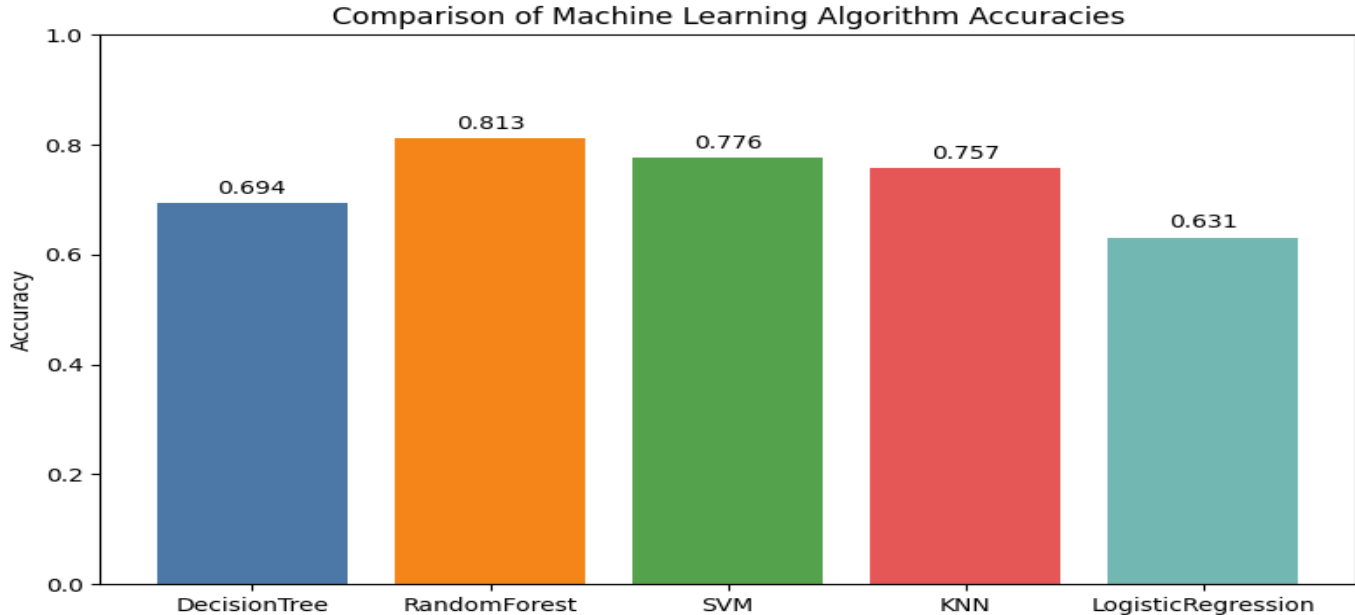
Preprocessing and Feature Engineering

Preprocessing included standardizing image sizes, applying Atmospheric Correction, and Image Normalization. Feature Engineering involved calculating vegetation indices, such as the Normalized Difference Vegetation Index (NDVI). Data Augmentation was performed using Rotation, Flips, and Cropping1. The data was divided into Training (70%), Validation (15%), and Test (15%) sets.

3. Traditional Machine Learning Experimentation (ML Baseline):

A comprehensive baseline was established using several traditional Machine Learning (ML) models before adopting Deep Learning.
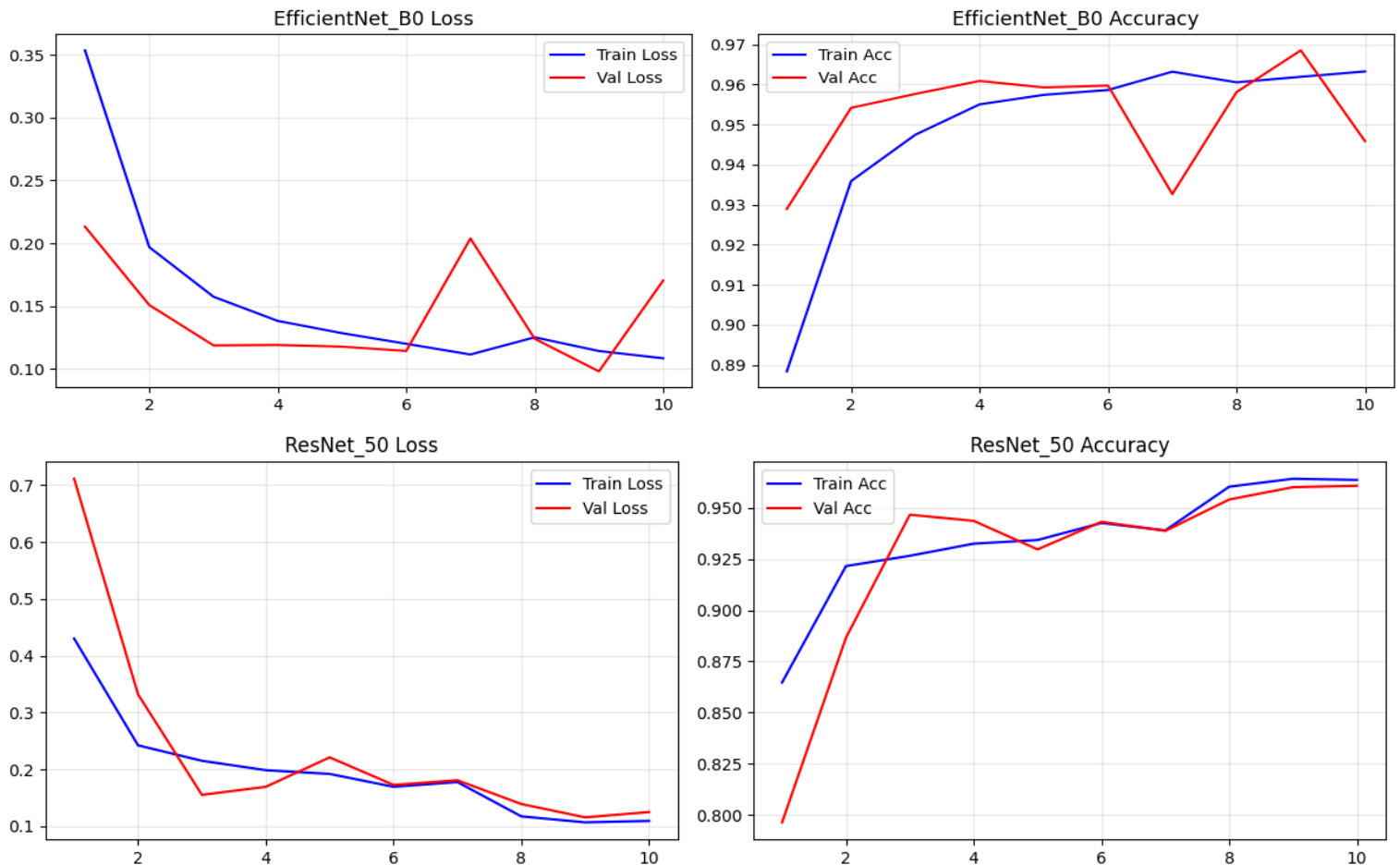
- **Models Tested:** The following algorithms were tested: **Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression**.
- **Input Data:** These models relied heavily on **manually engineered features** (e.g., NDVI and other spectral band statistics), rather than the raw pixel data, due to their limited capacity to handle the high dimensionality of raw multispectral images.

- **Conclusion:** While these models provided a quick performance benchmark, their reliance on manual feature extraction and limited ability to capture complex spatial patterns and spectral correlations inherent in Sentinel-2 imagery led to suboptimal performance compared to the later Deep Learning models. This necessitated the transition to CNNs for superior feature learning.

Comparison of Machine Learning Algorithm Accuracies

Algorithm	Accuracy
DecisionTree	0.694
RandomForest	0.813
SVM	0.776
KNN	0.757
LogisticRegression	0.631

4. Deep Learning Implementation and Model Selection

The transition to Deep Learning addressed the limitations of traditional ML by enabling **automatic feature extraction** from the multi-band imagery.

- **Architectures Tested:** Two advanced Convolutional Neural Network (CNN) architectures were tested: **ResNet-50** and **EfficientNet-B0**.
- **ResNet-50:** Chosen for its proven ability to train very deep networks by using skip connections, mitigating the vanishing gradient problem.
- **EfficientNet-B0 (Final Choice):** Selected for its balanced scaling of depth, width, and resolution. **EfficientNet-B0** achieved the **best overall performance** due to its superior efficiency, offering high accuracy while maintaining a relatively low computational footprint, making it ideal for scalable deployment.
- **Training Strategy:** Techniques like **Transfer Learning** (using ImageNet pre-trained weights) and **Fine-tuning** were applied. Training utilized **PyTorch or TensorFlow** with methods like **Cross-Validation** and **Early Stopping**.
- **Evaluation:** Performance was measured using **Accuracy, Precision, Recall, and F1-Score**. The final model demonstrated strong generalization capabilities on the Test Set.

5. *Model Deployment and Monitoring (MLOps)*

The finalized **EfficientNet-B0** model was prepared for a production environment.
- **Deployment:** The model was deployed via an API using frameworks like **Flask or FastAPI** and hosted on cloud platforms (Google Cloud, AWS, or Azure) to ensure scalability.
- **User Interface (UI):** A simple application was created to accept satellite images and provide visual classification results.
- **Monitoring:** Continuous monitoring was implemented to track performance, detect **Model Drift** , and set up alert systems for necessary intervention.
- **Retraining Strategy:** A plan for periodic retraining was established to incorporate new data and maintain accuracy over time.

6. *Conclusion and Future Recommendations*

The project successfully delivered a robust and efficient land classification system based on Sentinel-2 data, with **EfficientNet-B0** identified as the optimal classification engine. The comprehensive methodology, ranging from detailed preprocessing to advanced model deployment, ensures the system is production-ready.

Future enhancements include:
- Integrating data from other satellites like **Landsat**.
- Exploring newer algorithms such as **Transformers** for image classification.
- Further optimizing the deployment capabilities for greater real-world scalability.