Evaluation Report
Land Type Classification Using EfficientNetB0 on the EuroSAT Dataset

Contents

Chapter 1

Evaluation Overview

The trained deep learning model was evaluated on an unseen test subset of the EuroSAT RGB dataset to assess its generalization capability and classification performance.

The evaluation focused on standard multi-class classification metrics including accuracy, precision, recall, F1-score, and confusion matrix analysis.

Chapter 2

Dataset and Test Setup

The evaluation dataset consisted of RGB satellite image tiles from the EuroSAT dataset representing ten land-cover classes.

The original image resolution of $64 \times 64 \times 3$ was resized to $128 \times 128 \times 3$ during preprocessing to facilitate improved feature extraction within the EfficientNetB0 architecture.

2.1 Test Dataset Details

- Dataset: EuroSAT RGB
- Image resolution used: $128 \times 128 \times 3$
- Total test samples: 2016 images
- Number of classes: 10

2.2 Land-Cover Classes

- AnnualCrop
- Forest
- HerbaceousVegetation
- Highway

- Industrial
- Pasture
- PermanentCrop
- Residential
- River
- SeaLake

Chapter 3

Model Configuration

3.1 Base Architecture

The evaluation employed the EfficientNetB0 model pre-trained on the ImageNet dataset using a transfer learning approach.

3.2 Training Strategy

Two-stage training was performed:

- Initial training phase with all EfficientNetB0 backbone layers frozen.
- Fine-tuning phase with partial layer unfreezing.

3.3 Fine-Tuning Configuration

- Frozen backbone layers: First 220 layers
- Trainable backbone layers: Remaining upper layers
- Initial training epochs: 8
- Fine-tuning epochs: 10
- Best observed validation accuracy: 94.21%

Chapter 4

Hyperparameter Tuning

Short grid-style experiments were conducted to determine effective training parameters.

4.1 Tested Configurations

Configuration	Learning Rate	Dropout Rate
1	1×10^{-3}	0.3
2	5×10^{-4}	0.4
3	1×10^{-4}	0.5

Table 4.1: Tested hyperparameter configurations

The best performing setup used a learning rate of 1×10^{-3} with a dropout rate of 0.3 for the classifier head, followed by fine-tuning with a reduced learning rate of 1×10^{-5}.

Chapter 5

Overall Test Performance

5.1 Aggregate Metrics

Metric	Value
Test Loss	0.1623
Test Accuracy	94.44%
Macro Precision	95.27%
Macro Recall	93.85%
Macro F1-score	94.23%
Weighted F1-score	94.45%

Table 5.1: Overall classification performance metrics on the test dataset

These results demonstrate strong classification capability with balanced precision and recall across all classes.

Chapter 6

Per-Class Performance

6.1 Precision, Recall, and F1-score per Class

Class	Precision	Recall	F1-score	Support
AnnualCrop	0.92	0.94	0.93	220
Forest	0.95	0.98	0.96	212
HerbaceousVegetation	0.93	0.95	0.94	209
Highway	0.88	0.93	0.90	202
Industrial	0.98	0.95	0.96	195
Pasture	0.93	0.88	0.90	130
PermanentCrop	0.93	0.94	0.93	196
Residential	0.97	1.00	0.98	205
River	0.94	0.91	0.93	202
SeaLake	1.00	0.96	0.98	245

Table 6.1: Per-class precision, recall, F1-score, and support

6.2 Per-Class Accuracy

Class	Accuracy (%)
AnnualCrop	93.64
Forest	97.64
HerbaceousVegetation	94.74
Highway	92.57
Industrial	94.87
Pasture	87.69
PermanentCrop	93.88
Residential	99.51
River	91.09
SeaLake	95.92

Table 6.2: Per-class classification accuracy

Chapter 7

Confusion Matrix Analysis

The confusion matrix exhibited strong diagonal dominance, indicating accurate predictions across classes.

Key observations include:

- Residential achieved the highest accuracy (99.51%) with minimal misclassification.
- Forest and SeaLake also demonstrated excellent performance with very low class overlap.
- Pasture showed the lowest accuracy (87.69%), mainly confused with vegetation-related classes such as AnnualCrop and HerbaceousVegetation.
- Highway and River categories demonstrated moderate confusion due to similar surrounding textures.
- Industrial regions were well distinguished with consistently strong precision and recall.

Chapter 8

Fine-Tuning Impact

The peak validation accuracy of 94.21% occurred during the initial transfer learning phase prior to fine-tuning.

Further partial unfreezing of the backbone did not produce additional validation improvement across the limited epochs tested. However, fine-tuning maintained stable performance without signs of overfitting, confirming the robustness of the model representations.

Chapter 9

Conclusion

The optimized EfficientNetB0-based classifier demonstrated strong generalization across diverse land-cover categories.

Key outcomes:

- Overall test accuracy of 94.44%.
- Macro F1-score of 94.23% showing balanced performance across classes.
- Exceptional recognition of urban and waterbody categories.
- Minor confusion observed primarily among vegetation-based classes.

These results confirm that transfer learning using EfficientNetB0 is highly effective for satellite land-cover classification and fulfills the performance targets established in Milestone 3.