

Research Methodology Overview

PHASE 1
DATA

EuroSAT

27,000 images • 10 classes
64×64 px • Sentinel-2 (10 m)

UC Merced

2,100 images • 21 classes
256×256 px • USGS Aerial (0.3 m)

PHASE 2
PREPROCESS

Data Preprocessing

Resize to 224×224 • Stratified 80/20 split (seed=42) • RGB channels only

Data Augmentation

Random H/V flip • Rotation ($\pm 15^\circ$) • Color jitter • ImageNet normalization

8 Model Architectures (3 Families) — All ImageNet Pretrained

PHASE 3
MODELING

Classical CNN

ResNet-50 (23.5M)
ResNet-101 (42.5M)
DenseNet-121 (7.0M)
EfficientNet-B0 (4.0M)
EfficientNet-B3 (10.7M)

Vision Transformer

ViT-B/16 (85.8M)
Swin-T (27.5M)

Modern CNN

ConvNeXt-Tiny (27.8M)

PHASE 4
EVALUATION

Uniform Training Protocol

AdamW ($lr=10^{-4}$) • ReduceLROnPlateau • Early stopping (patience=10) • 30 epochs • Batch 32

Performance Evaluation

Overall Accuracy • F1-Macro • F1-Weighted • Cohen's κ • Per-class Precision / Recall / F1

Statistical Testing

McNemar's test
(pairwise significance)

Error Analysis

Confusion matrices
Misclassified samples

Efficiency Analysis

Parameters vs. accuracy
Training time

PHASE 5
FINDINGS

Key Findings

- Architecture matters less than training recipe with transfer learning
- ConvNeXt-T best on EuroSAT (99.06%) • EfficientNet-B3 best on UC Merced (99.76%)
- Most pairwise differences not statistically significant (McNemar's $p > 0.05$)
- EfficientNet-B0 (4.0M params) within 1% of all models — best efficiency