Comparative Analysis of CNN Architectures for Land Cover Classification on the EuroSAT Dataset

Ashwin Mankar
23b0726
23b0726@iitb.ac.in

Karan Satarkar
23b0708
23b0708@iitb.ac.in

***Abstract*—This project evaluates multiple convolutional neural network (CNN) architectures for land cover classification on the EuroSAT dataset. Six representative models were selected across three architecture families: ResNet (ResNet18, ResNet50), DenseNet (DenseNet121, DenseNet201), and EfficientNet (EfficientNet-B0, EfficientNet-B4). A unified training pipeline using PyTorch, AdamW optimizer, StepLR learning rate schedule, and RandAugment data augmentation was implemented. We report training dynamics, accuracy curves, per-epoch metrics, and comparative performance across the models. Results show significant performance differences between lightweight and standard variants, demonstrating the impact of depth, parameter count, and architectural design on remote sensing classification accuracy.**

***Index Terms*—EuroSAT, CNN, ResNet, DenseNet, EfficientNet, Remote Sensing, PyTorch.**

I. Introduction

Deep learning has shown exceptional performance on remote sensing tasks, particularly land-cover classification. The EuroSAT dataset provides 27,000 labeled Sentinel-2 images across 10 classes. This project evaluates and compares multiple CNN architectures under a unified training framework to identify trade-offs between accuracy, model size, and computational cost.

II. Dataset

EuroSAT consists of 64×64 RGB satellite images belonging to 10 land cover categories such as Forest, Residential, Pasture, PermanentCrop, and others. We divided the dataset into fixed splits: 70% training, 15% validation, and 15% testing.

All images were resized to 64×64. The same normalization statistics were used across all models for fair comparison. Heavy augmentation (RandAugment) was applied to the training set.

III. Models Evaluated

We selected two representative models from each family:

- **ResNet:** ResNet18 (lightweight), ResNet50 (standard).
- **DenseNet:** DenseNet121 (lightweight), DenseNet201 (standard).
- **EfficientNet:** EfficientNet-B0 (lightweight), EfficientNet-B4 (standard).

All models were implemented via the TIMM library and trained using a unified PyTorch pipeline.

IV. Model Architectures

This work evaluates three major CNN architecture families: ResNet, DenseNet, and EfficientNet. Each family includes a lightweight and a standard variant to compare the effect of model depth and capacity on performance.

ResNet (Residual Networks): ResNet introduces residual connections that allow gradients to flow through shortcut paths, enabling stable training of deep networks. The lightweight ResNet18 contains fewer layers and parameters, while the deeper ResNet50 uses bottleneck residual blocks to improve feature extraction capability. The skip connections in both variants prevent vanishing gradients and enable effective training even at greater depths.

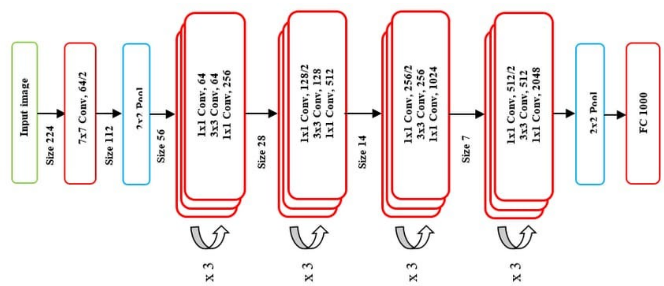

Fig. 1. ResNet architecture illustration.

DenseNet (Densely Connected Networks): DenseNet connects each layer to every subsequent layer within the same block, ensuring feature reuse and efficient gradient flow. DenseNet121 is the lightweight version with fewer dense blocks, while DenseNet201 increases depth and growth rate for richer feature maps. This dense connectivity reduces redundancy and encourages compact feature representations.

EfficientNet: EfficientNet scales depth, width, and resolution in a balanced manner using compound scaling. EfficientNet-B0 serves as the lightweight baseline, whereas EfficientNet-B4 expands resolution and channel width for higher accuracy. These models rely on inverted residual blocks and squeeze–and–excitation (SE) layers to achieve strong accuracy–efficiency trade-offs.

Across all three families, the lightweight variants prioritize computational efficiency, while the standard variants focus on performance through larger capacity and deeper feature hierarchies.

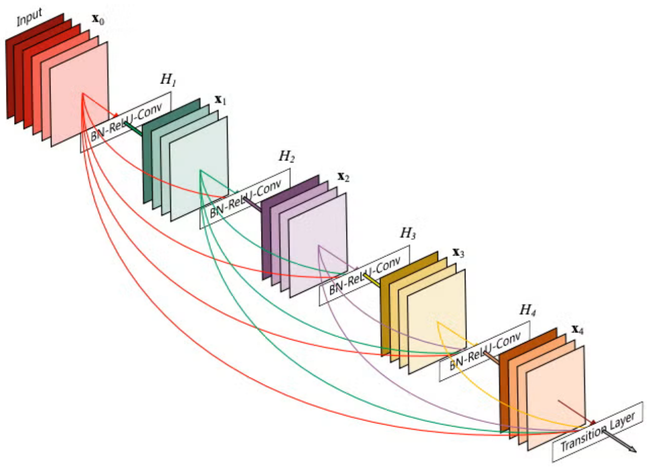

Fig. 2. DenseNet architecture illustration.

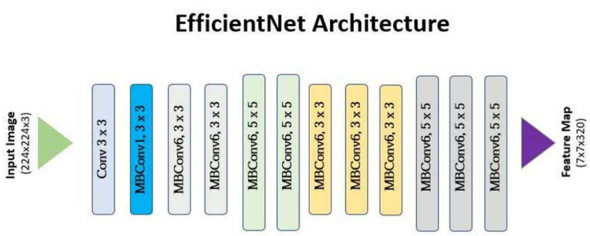

Fig. 3. EfficientNet architecture illustration.

V. Methodology

A. Training Setup

We used:

- Optimizer: AdamW
- Learning rate: 1e-3
- Scheduler: StepLR
- Batch size: 64 for small models; 32 for heavier ones
- Epochs: 20

The training loop logs per-epoch loss, accuracy, and saves these plots and also the best model.

B. Evaluation Metrics

We measure:

- Training and validation accuracy
- Training and validation loss
- Test accuracy on normal and noisy images
- Parameter count and Time taken for training

VI. Results

A. Validation Accuracy Curves

Figure 5 shows the overlay plot of validation accuracy vs epoch for all six models.

B. Validation Loss Curves

Figure 5 shows the overlay plot of validation loss vs epoch for all six models.

C. Final Accuracy and Test Performance Comparison

Table I summarizes the final validation accuracies of all six models, while Figure 6 shows the corresponding test accuracies on the clean test set. Both metrics show a consistent trend across architectures.

TABLE I
FINAL VALIDATION ACCURACY OF ALL MODELS

Model	Parameters	Val Accuracy
ResNet18	11181642	0.980494
ResNet50	23528522	0.982469
DenseNet121	6964106	0.980000
DenseNet201	18112138	0.980494
EfficientNet-B0	4020358	0.979012
EfficientNet-B4	17566546	0.974568

Across both validation and test evaluations, **ResNet50 consistently emerges as the strongest performer**. It achieves the highest validation accuracy (0.982) and also tops the clean test accuracy with 0.981. EfficientNet-B0 and DenseNet201 follow closely behind, while EfficientNet-B4 shows the lowest performance among the six. The clear margin by which ResNet50 leads indicates its strong generalization ability and stable feature extraction, making it the most reliable model in this benchmark.

D. Robustness Analysis

To evaluate model reliability under real-world degradations, we conducted four robustness tests by adding controlled perturbations to the test set. The evaluated corruption types were: occlusion, noise, brightness shift, and blur. For each corruption, accuracy was measured across increasing severity levels.

E. Robustness Analysis

To evaluate model reliability under real-world degradations, we conducted four robustness tests by adding controlled perturbations to the test set. The evaluated corruption types were: occlusion, noise, brightness shift, and blur. For each corruption, accuracy was measured across increasing severity levels.

Across all corruption types, ResNet50 demonstrates the strongest overall robustness. It consistently outperforms the other five models under occlusion and blur, maintaining the highest accuracy even at severe corruption levels. Although it does not rank first in noise and brightness degradations, it remains within the top two or three models, showing minimal accuracy drop compared to others. These results indicate that ResNet50's architecture preserves more stable feature representations under perturbations, making it the most reliable model when inputs are degraded.

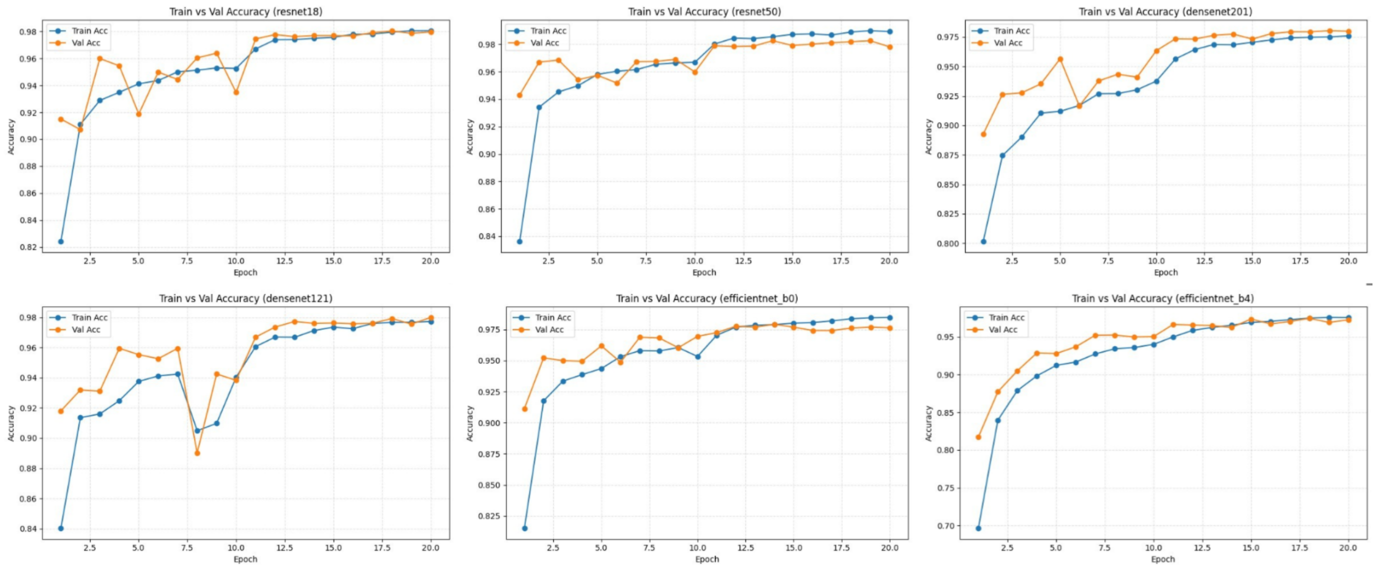

Fig. 4. Validation accuracy comparison across all models.

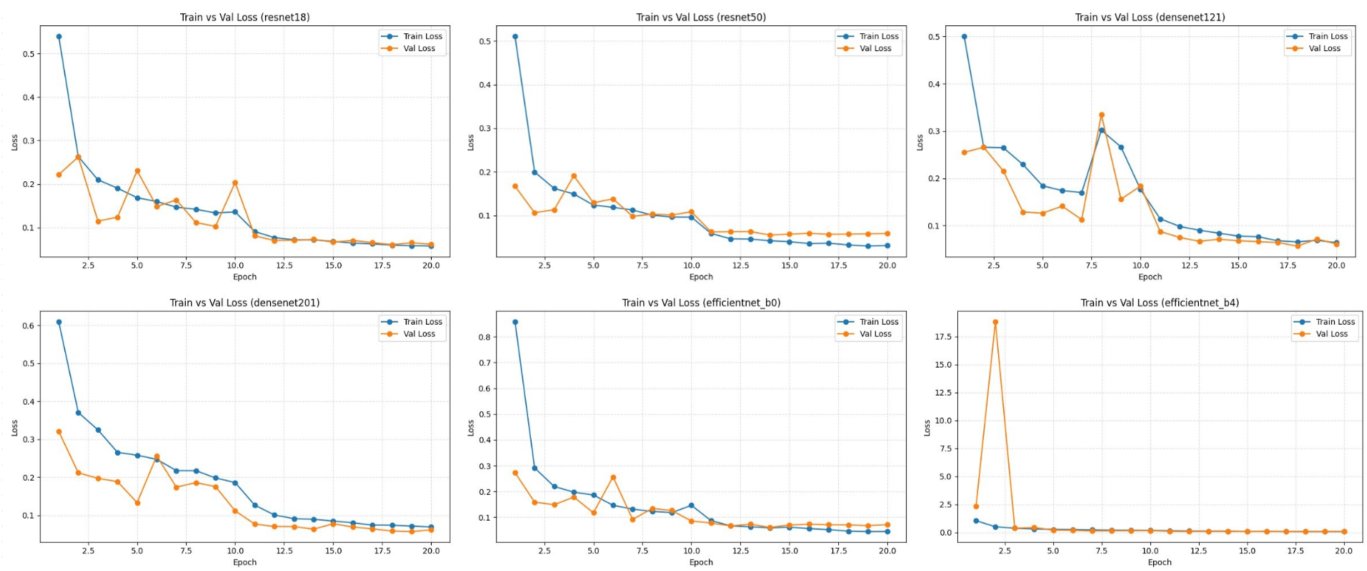

Fig. 5. Validation accuracy comparison across all models.

VII. Discussion

The training and validation curves across all six models show stable learning behaviour, with no signs of divergence or oscillation. All models exhibit smooth convergence, indicating that the unified training pipeline, AdamW optimizer, and StepLR scheduler provided consistent gradient flow and optimization stability.

A common pattern across the architectures is the early plateau observed within the first 5–7 epochs. Most models achieve rapid accuracy gains during the initial epochs, after which improvement slows significantly. This suggests that the networks quickly learn low-level features and spend the remaining epochs refining class-specific patterns. The plateau is most evident in ResNet50 and DenseNet201, both of which maintain high accuracy with minimal variance in later epochs.

The use of RandAugment contributes notably to improved generalization. Across all models, the gap between training and validation accuracy remains small, and validation curves do not show overfitting spikes. Lightweight architectures such as EfficientNet-B0 and DenseNet121 particularly benefit from these augmentations, achieving performance close to heavier models despite having fewer parameters.

We also generated confusion matrices for all models to inspect patterns of misclassification. However, since the overall error rate was extremely low across classes, the matrices did not reveal any meaningful trends or systematic confusions; most cells were near-zero, reflecting the high accuracy already

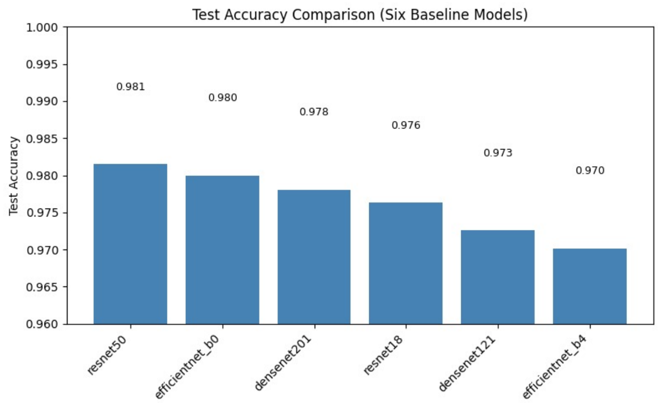

Fig. 6. Test accuracy comparison across all six models.

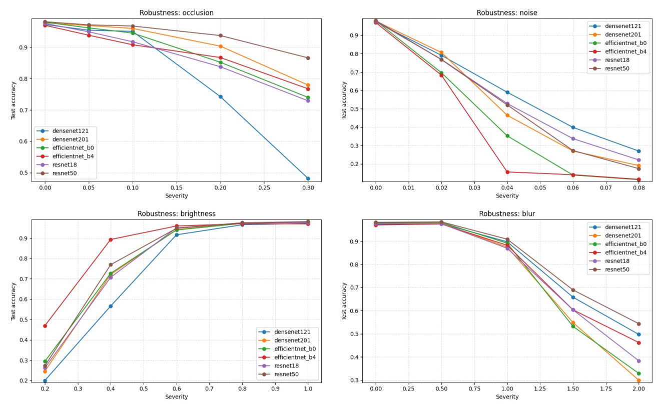

Fig. 7. Robustness evaluation across four corruption types: occlusion, noise, brightness shift, and blur.

observed in the quantitative metrics.

A clear model capacity versus accuracy trend emerges. While deeper or wider models generally perform better, the improvement is not strictly proportional to parameter count. ResNet50 achieves the highest validation and test accuracy despite not being the largest model. In contrast, EfficientNet-B4, one of the heavier networks, underperforms several smaller architectures. This highlights that, for EuroSAT's resolution and dataset size, excessively large models may not fully exploit their capacity, whereas architectures with balanced depth and connectivity—such as ResNet50—achieve the best trade-off between accuracy and computational cost.

VIII. Conclusion

This project demonstrates that CNN architecture design, depth, and feature connectivity patterns significantly influence performance on the EuroSAT land-cover classification task. While deeper models generally achieve higher accuracy, the improvement does not always scale with parameter count, as seen with EfficientNet-B4, which underperforms several lighter models.

Across all evaluations, ResNet50 consistently emerges as the strongest architecture. It achieves the highest validation and clean test accuracy, shows stable training behaviour, and maintains superior robustness under occlusion and blur. Models such as ResNet18 and EfficientNet-B0 offer competitive performance with substantially lower computational cost, making them viable for deployment in resource-limited settings.

Overall, ResNet50 provides the best balance of accuracy, stability, and robustness, making it the most reliable model for practical land-cover classification applications based on the EuroSAT dataset.

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[2] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 4700–4708.

[3] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. 36th Int. Conf. Machine Learning (ICML)*, 2019, pp. 6105–6114.