

Track 1 – Image Classification Report

Category Compliance

Category 1 — Pretrained model weights were used. We fine-tuned EfficientNet-B3 pretrained on ImageNet.

Team Name: GSA

Team members: Gautam Balgi, Siddhant Kasture, Ameya Bhalerao

1. Preprocessing and Augmentation

All images were resized to 224×224 pixels. During training the following augmentations were applied to improve generalization on the small dataset:

- Random resized crop (scale 0.6–1.0)
- Random horizontal and vertical flips
- Color jitter (brightness, contrast, saturation, hue)
- Random rotation ($\pm 30^\circ$)
- Random grayscale (5% probability)
- Random perspective distortion
- Random erasing (20% probability)

Validation and test images were resized to 246×246 and center cropped to 224×224 with standard ImageNet normalization applied to all splits.

2. Model Architecture

We used EfficientNet-B3 pretrained on ImageNet as the backbone. The original classifier was replaced with a custom head:

Dropout(0.5) → Linear(1536 → 512) → ReLU → Dropout(0.3) → Linear(512 → 102)

This two-layer head gave better results than a single linear layer, likely due to the added capacity for learning class-specific features.

3. Training Procedure and Hyperparameters

Training was done in two phases to prevent early destruction of pretrained features:

Phase 1 (epochs 1–5): Backbone frozen, only classifier trained with LR = $1e-3$

Phase 2 (epoch 6 onwards): Full model fine-tuned with LR = $3e-4$

Additional details:

- Optimizer: AdamW (weight decay = $1e-4$)
- Loss: Cross-entropy with label smoothing (0.1) and class weights for Macro F1
- Scheduler: Cosine annealing (min LR = $1e-6$)
- Batch size: 16
- Early stopping: patience of 8 epochs on validation accuracy
- Best validation accuracy: 0.90 (epoch 10)

Class weights were computed using sklearn's balanced class weight method to address class imbalance and improve Macro F1 on minority classes.

4. Inference

At inference time, Test Time Augmentation (TTA) was applied with 5 transforms per image:

1. Original center crop
2. Horizontal flip
3. Vertical flip
4. Larger scale crop ($1.2\times$)
5. Random crop

Logits from all 5 passes were averaged before taking the final argmax prediction.

Validation Macro F1: 0.9135