

**EECE 693 - WiDS Hackathon Report**

Submitted by

Lynn Ariss, Tamara Fakih, Dana Kossaybati, Omar Ramadan

Term: Fall 2025 – 2026

Date of Submission: Tuesday, October 28th, 2025

**MobileNetV3‑Small / EfficientNet‑B0 — Training Setup**

To complement the YOLOv8‑cls baseline, we trained lightweight ImageNet‑pretrained classifiers using torchvision—MobileNetV3‑Small and EfficientNet‑B0—on the same UXORD‑10K folder layout. The pipeline kept the dataset interface identical across models (train/val/test with class subfolders) but raised the input resolution to 320×320 to recover fine‑grained cues on small munitions. We applied standard ImageNet normalization and a moderately strong augmentation stack (RandomResizedCrop, horizontal/vertical flips, ColorJitter, a light AutoAugment policy, and RandomErasing). Optimization used AdamW (learning rate 1e‑3, weight decay 5e‑4) with label smoothing at 0.1 and a CosineAnnealingLR scheduler across 40 epochs. To stabilize early training on small backbones, we adopted a 5‑epoch warmup with the feature extractor frozen and then unfroze the backbone for full fine‑tuning. Batch size was auto‑selected up to the largest value that fit GPU memory (128 on a Tesla T4). Checkpoints were saved whenever the validation Top‑1 improved, and we enforced early stopping on a patience window of ten epochs. This setup yields a fair, apples‑to‑apples comparison with YOLOv8‑cls while testing whether mobile‑class backbones can deliver competitive accuracy with faster, lighter inference for fieldable deployments.

**MobileNetV3‑Small — Test Results and Visual Analysis**

Using the best checkpoint selected by validation Top‑1, MobileNetV3‑Small achieved a Top‑1 accuracy of 0.7355 and a Top‑5 accuracy of 0.9466 on the held‑out test split (n=431). The sklearn cross‑check matched those figures exactly, and the evaluation exported a per‑class CSV, confusion matrix, per‑class precision/recall/F1 bars, and a mixed prediction grid for qualitative review. Per‑class behavior is instructive: aircraft‑bombs and projectiles posted strong F1 scores (≈0.83 and ≈0.81), while fuzes and mortars showed high precision (≈0.89 and ≈0.91) coupled with lower recall (≈0.76 and ≈0.70), suggesting conservative decisions that miss some positives. Landmines exhibited very high recall (≈0.91) but modest precision (≈0.61), meaning the model captures most true instances at the cost of more false alarms—acceptable in safety‑first triage but worth tightening. The most challenging categories were rockets and submunitions (F1 ≈0.63 and ≈0.55, recalls ≈0.56 and ≈0.48), where shape similarity and scene clutter likely drive confusions. Visually, the confusion matrix concentrates errors among small, visually similar munitions and between elongated forms such as mortars and projectiles. These observations point to three levers for improvement: higher input resolution or multi‑crop evaluation to capture small parts, modest architecture scaling (e.g., EfficientNet‑B0 or an ensemble) to raise Top‑1 without sacrificing speed, and data‑centric iterations that emphasize hard negatives, class‑aware sampling, and augmentation tuned to field conditions. Notably, the MobileNetV3‑Small results exceed the YOLOv8‑cls Top‑1 by roughly four to five percentage points on this test set while keeping Top‑5 above 94%, reinforcing the viability of mobile‑class models for low‑compute deployments where latency, battery, and offline operation matter.