# Intern Brief: Evaluating D-2 Pretraining Benefit

* Maintain parallel `.docx` and `.md` files for every write-up; keep English versions strictly in English.
* Re-run the cross-dataset transfer pipeline (`python scripts/experiments/pretrain\_transfer.py`) focusing on the `D-1 + D-3 → D-2` scenario while sweeping five distinct seeds (e.g., 42, 43, 44, 45, 46) at a fixed D2 fine-tune ratio of 0.4.
* For each seed, record two configurations that share the identical D2 fine-tune split: (a) pretrain on D-1 + D-3 then fine-tune on 40% of D2; (b) train only on the same 40% slice of D2 without pretraining.
* Ensure both configurations reuse the exact D2 train/validation/test partitions so the comparison isolates pretraining; log the split seeds you used.
* Append the new runs to `results/domain\_adaptation\_default.csv`, tagging rows clearly (`mode=pretrain\_finetune` vs. `mode=target\_only`).
* Build comparison tables or plots for AUROC, Accuracy, and AUPRC across seeds, highlighting consistent gains or regressions.
* Review prior notes in `PROGRESS.md` and existing CSV outputs to judge whether pretraining shows a stable benefit pattern; call out scenarios where it does not.
* Summarize methods, metrics, and takeaways (including whether pretraining is practically meaningful) and capture them in both language tracks once analysis is complete.