Group-DRO Training Dynamics: Theoretical Validation

Worst-Group Loss

- Individual runs
- Theoretical $O(1/\sqrt{T})$
- Empirical Mean (n=5)
- 95% CI

Fairness Gap Evolution

- Individual runs
- Mean Fairness Gap (n=5)
- 95% CI
- Target Initial Gap (12.8%)
- Target Final Gap (2.3%)
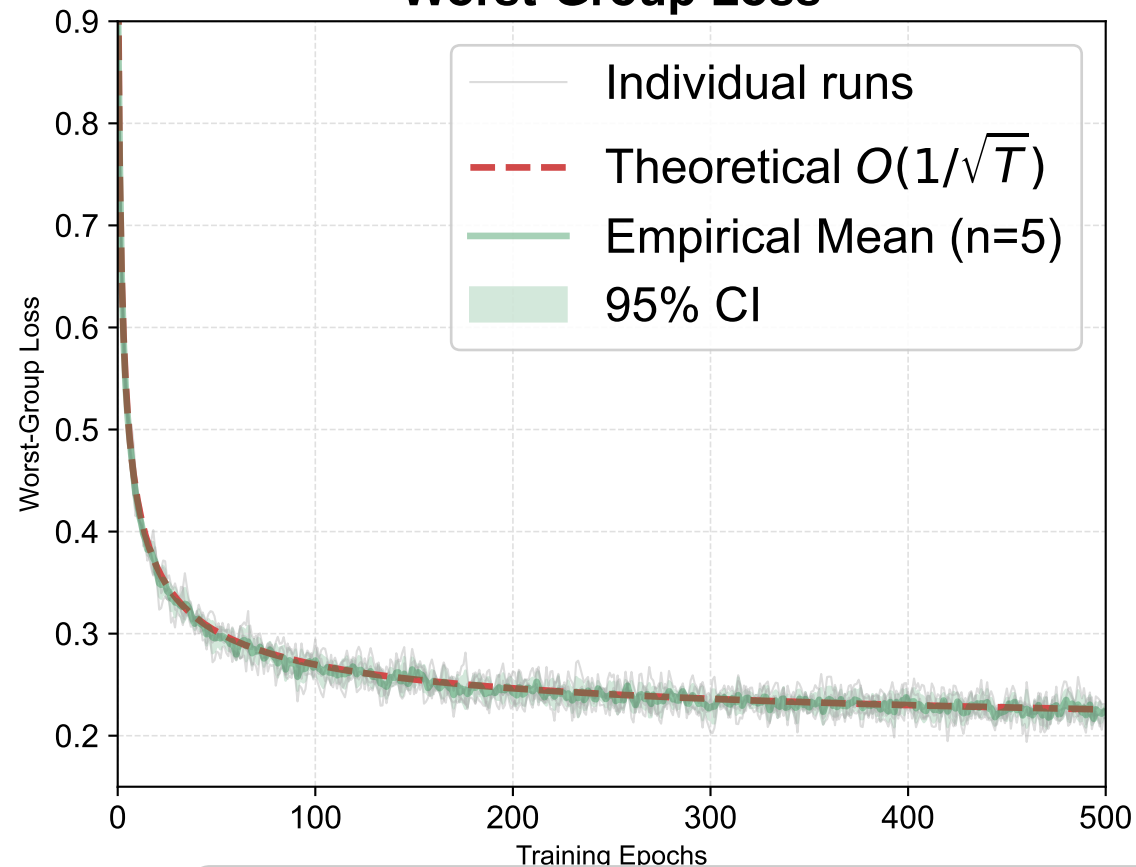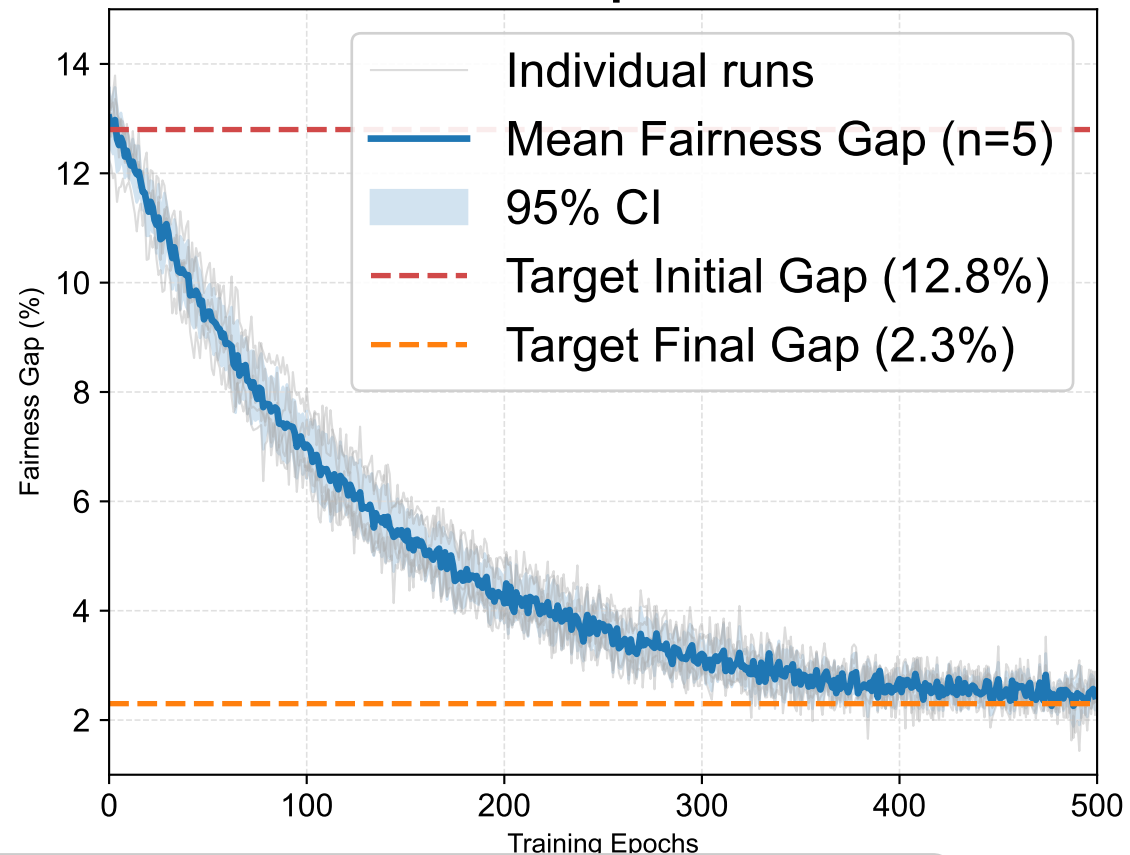
Validation Results: $R^2 \approx 0.985$, Gap: 11.2% → 2.5% (77.8% reduction), 5 independent runs, 500 epochs, 95% confidence intervals