Accuracy Variability Across Seeds (3 seeds per experiment)