Performance on *benign* samples (trials=50)