Distribution of test accuracy in batch (no. architectures 300)