racy during training process for different data sizes, numbers of parameters, and numbers of layers

Experiment 3:
model size, number of layers
- 16, 1
- 32, 1
- 64, 1
- 128, 1
- 16, 2
- 32, 2
- 64, 2
- 128, 2

data size
- 1000
- 10000
- 50000
- 100000