Mean during training process for different data sizes, activation functions, and number of layers