Unterschied der Modelle ist insignifikant.

Jedoch Trainingszeit bei 3 layern mehr als Doppelt so hoch.

Lässt vermuten, dass dieses Problemstellung nicht mehr Komplexität benötigt