ID	epochs	learn_rate	batch_size	val_split	verbose	Network Shape	Test error	NOTES
0-0	500	0.001	40	0.2	1	[X],2,2,1	0.0511	
0-1	500	0.001	40	0.2	1	[X],2,2,1	0.0403	
0-2	500	0.001	40	0.2	1	[X],2,2,1	0.0404	
0-3	500	0.001	40	0.2	1	[X],2,2,1	0.0553	
0-4	500	0.001	40	0.2	1	[X],2,2,1	0.0637	
0-5	500	0.001	40	0.2	1	[X],2,2,1	0.0484	
0-6	500	0.001	40	0.2	1	[X],2,2,1	0.0381	
0-7	500	0.001	40	0.2	1	[X],2,2,1	0.0791	
0-8	500	0.001	40	0.2	1	[X],2,2,1	0.0644	
0-9	500	0.001	40	0.2	1	[X],2,2,1	0.0523	
mean							0.05331	Baseline, converges ~350 epochs
1-0	600	0.001	40	0.2	1	[X],2,2,1	0.0397	
1-1	600	0.001	40	0.2	1	[X],2,2,1	0.0626	
1-2	600	0.001	40	0.2	1	[X],2,2,1	0.0588	
1-3	600	0.001	40	0.2	1	[X],2,2,1	0.0482	
1-4	600	0.001	40	0.2	1	[X],2,2,1	0.0559	
mean							0.05304	Not much change.
2-0	500	0.003	40	0.2	1	[X],2,2,1	0.0609	
2-1	500	0.003	40	0.2	1	[X],2,2,1	0.0619	
2-2	500	0.003	40	0.2	1	[X],2,2,1	0.0624	
2-3	500	0.003	40	0.2	1	[X],2,2,1	0.0454	
2-4	500	0.003	40	0.2	1	[X],2,2,1	0.0478	
mean							0.05568	Converges ~150 epochs, but increased MSE. Potential loss of minutiae in data
3-0	500	0.001	60	0.2	1	[X],2,2,1	0.0595	
3-1	500	0.001	60	0.2	1	[X],2,2,1	0.0389	
3-2	500	0.001	60	0.2	1	[X],2,2,1	0.0616	
3-3	500	0.001	60	0.2	1	[X],2,2,1	0.0599	
3-4	500	0.001	60	0.2	1	[X],2,2,1	0.0583	
mean							0.05564	Earlier convergence at ~250 epochs, more accurate gradient est.
4-0	500	0.001	40	0.2	1	[X],2,2,2,1	0.1969	

4-1	500	0.001	40	0.2	1	[X],2,2,2,1	0.0738	
4-2	500	0.001	40	0.2	1	[X],2,2,2,1	0.0435	
4-3	500	0.001	40	0.2	1	[X],2,2,2,1	0.0644	
4-4	500	0.001	40	0.2	1	[X],2,2,2,1	0.0542	
mean							0.08656	Added layer causes more complex gradient, model hasn't converged by 500 epochs
5-0	700	0.001	40	0.2	1	[X],2,2,2,1	0.0386	
5-1	700	0.001	40	0.2	1	[X],2,2,2,1	0.0429	
5-2	700	0.001	40	0.2	1	[X],2,2,2,1	0.0428	
5-3	700	0.001	40	0.2	1	[X],2,2,2,1	0.0371	
5-4	700	0.001	40	0.2	1	[X],2,2,2,1	0.0367	
mean							0.03962	With more epochs, convergence occurs around ~550 epochs
6-0	700	0.001	60	0.2	1	[X],2,2,2,1	0.055	
6-1	700	0.001	60	0.2	1	[X],2,2,2,1	0.0476	
6-2	700	0.001	60	0.2	1	[X],2,2,2,1	0.0358	
6-3	700	0.001	60	0.2	1	[X],2,2,2,1	0.0378	
6-4	700	0.001	60	0.2	1	[X],2,2,2,1	0.0421	
mean							0.04366	Earlier convergence ~450 epochs. Slight increase in MSE.
7-0	700	0.001	80	0.2	1	[X],2,2,2,1	0.0295	
7-1	700	0.001	80	0.2	1	[X],2,2,2,1	0.0287	
7-2	700	0.001	80	0.2	1	[X],2,2,2,1	0.0342	
7-3	700	0.001	80	0.2	1	[X],2,2,2,1	0.0309	
7-4	700	0.001	80	0.2	1	[X],2,2,2,1	0.0291	
mean							0.03048	Higher batch count is able to better capture the gradient, and earlier. ~400 epochs.
8-0	700	0.001	80	0.2	1	[X],3,2,2,1	0.0314	
8-1	700	0.001	80	0.2	1	[X],3,2,2,1	0.0305	
8-2	700	0.001	80	0.2	1	[X],3,2,2,1	0.0292	
8-3	700	0.001	80	0.2	1	[X],3,2,2,1	0.0294	
8-4	700	0.001	80	0.2	1	[X],3,2,2,1	0.0295	
mean							0.03	Added neuron helps capture more information earlier (converges ~300 epochs)
9-0	700	0.001	80	0.2	1	[X],5,3,2,1	0.0307	

9-1	700	0.001	80	0.2	1	[X],5,3,2,1	0.0319	
9-2	700	0.001	80	0.2	1	[X],5,3,2,1	0.0314	
9-3	700	0.001	80	0.2	1	[X],5,3,2,1	0.0316	
9-4	700	0.001	80	0.2	1	[X],5,3,2,1	0.0306	
mean							0.03124	More consistent performance across tests without much gain in error.