Reflection Report

In this experiment, I started with a neural network consisting of four layers: one input layer, two hidden layers, and one output layer. This initial model had 416,962 trainable parameters and took approximately 108.86 seconds to train.

To attempt improvement, I included an additional hidden layer of 8 neurons. This reduced the number of trainable parameters to 416,954 and the training time to 119.02 seconds. Finally, I built a more complex variant by introducing two hidden layers with 64 and 8 neurons, which placed the parameter count at 835,162 and the training time at 141.33 seconds.

From the training results, the deeper layer networks maintained or even improved a little bit. For example, the 4 hidden layer model achieved high precision and recall on many classes, showing stronger learning capacity.

Yet on the test set, all models struggled with generalizing. There were numerous classes with low recall and F1-scores, regardless of model complexity. In particular, whereas the 4-layer model performed exceptionally well while training, it did not achieve notable improvement on the test data and even showed signs of overfitting in a few cases.

This observation is to highlight that additional model complexity does not directly lead to better test performance. It just emphasizes the requirement for techniques like regularization, dropout, or better feature engineering. In the future, I will look into such methods to promote generalization and avoid overfitting.

Table that summarizes the model results

Training Dataset

With 2 hidden layers					With 3 hidden layers					With 4 hidden layers			
precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
0.88	0.84	0.86	1561		0.8	0.89	0.84	1561		0.84	0.86	0.85	1561
0.88	0.76	0.82	648		0.95	0.7	0.8	648		0.85	0.77	0.81	648
0.82	0.75	0.78	571		0.76	0.78	0.77	571		0.83	0.73	0.78	571
0.98	0.99	0.99	434		0.99	0.98	0.99	434		0.99	0.99	0.99	434
0.95	0.91	0.93	620		0.91	0.93	0.92	620		0.9	0.94	0.92	620
0.98	0.95	0.96	174		0.96	0.96	0.96	174		0.98	0.95	0.97	174
0.94	0.91	0.92	381		0.94	0.89	0.92	381		0.97	0.88	0.92	381
0.74	0.91	0.82	1580		0.76	0.88	0.82	1580		0.86	0.79	0.82	1580
0.85	0.79	0.82	1334		0.9	0.75	0.82	1334		0.85	0.78	0.82	1334
0.88	0.85	0.86	401		0.91	0.83	0.87	401		0.85	0.83	0.84	401
0.8	0.7	0.75	149		0.83	0.68	0.75	149		0.75	0.74	0.75	149
0.88	0.93	0.9	932		0.9	0.9	0.9	932		0.96	0.86	0.9	932
0.86	0.84	0.85	2310		0.83	0.87	0.85	2310		0.77	0.91	0.84	2310
0.96	0.96	0.96	175		0.93	0.98	0.96	175		0.92	0.98	0.95	175
0.97	0.94	0.95	1004		0.96	0.95	0.95	1004		0.94	0.94	0.94	1004
0.93	0.86	0.9	428		0.94	0.86	0.89	428		0.8	0.88	0.84	428
0.88	0.94	0.91	372		0.91	0.93	0.92	372		0.88	0.92	0.9	372
0.81	0.86	0.83	585		0.9	0.79	0.85	585		0.87	0.78	0.82	585
0.87	0.87	0.87	0.87		0.86	0.86	0.86	0.86		0.86	0.86	0.86	0.86
0.89	0.87	0.88	13659		0.89	0.86	0.88	13659		0.88	0.86	0.87	13659
0.87	0.87	0.87	13659		0.87	0.86	0.86	13659		0.86	0.86	0.86	13659

Testing Dataset

With 2 hidden layers					With 3 hidden layers					With 4 hidden layers			
precision	recall	f1-score	support		precision	recall	f1-score	support		precision	recall	f1-score	support
0.54	0.21	0.3	617		0.37	0.24	0.29	617		0.13	0.79	0.23	617
0.39	0.12	0.18	273		0.47	0.1	0.16	273		0.35	0.12	0.18	273
0.39	0.19	0.25	243		0.18	0.27	0.21	243		0.39	0.18	0.25	243
0.93	0.18	0.3	214		0.95	0.18	0.3	214		0.93	0.18	0.3	214
0.6	0.17	0.27	258		0.52	0.17	0.26	258		0.53	0.17	0.26	258
0.89	0.15	0.25	54		0.67	0.15	0.24	54		0.89	0.15	0.25	54
0.69	0.18	0.29	171		0.67	0.16	0.26	171		0.67	0.16	0.26	171
0.44	0.31	0.36	664		0.26	0.38	0.31	664		0.52	0.22	0.31	664
0.46	0.2	0.28	569		0.5	0.2	0.28	569		0.47	0.2	0.28	569
0.54	0.25	0.34	199		0.62	0.24	0.35	199		0.58	0.25	0.35	199
0.33	0.09	0.15	85		0.33	0.08	0.13	85		0.23	0.09	0.13	85
0.62	0.22	0.32	399		0.69	0.2	0.31	399		0.75	0.19	0.3	399
0.19	0.77	0.31	968		0.21	0.65	0.32	968		0.51	0.33	0.4	968
0.39	0.11	0.17	82		0.36	0.11	0.17	82		0.38	0.12	0.19	82
0.83	0.23	0.36	472		0.77	0.23	0.35	472		0.71	0.23	0.35	472
0.59	0.17	0.26	188		0.54	0.16	0.25	188		0.42	0.17	0.24	188
0.61	0.3	0.4	152		0.67	0.3	0.42	152		0.68	0.3	0.41	152
0.42	0.2	0.27	246		0.54	0.19	0.28	246		0.48	0.19	0.27	246
0.3	0.3	0.3	0.3		0.3	0.3	0.3	0.3		0.28	0.28	0.28	0.28
0.55	0.22	0.28	5854		0.52	0.22	0.27	5854		0.53	0.22	0.28	5854
0.5	0.3	0.3	5854		0.46	0.3	0.29	5854		0.51	0.28	0.3	5854

Parameters and training time

Initial Model,

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 32)	416,128
dense_1 (Dense)	(None, 16)	528
dense_2 (Dense)	(None, 18)	306

Total params: 1,250,888 (4.77 MB)

Trainable params: 416,962 (1.59 MB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 833,926 (3.18 MB)

Training time: 108.86 seconds

Modified Model, with 3 hidden layers

Model: "sequential_1"

Layer (type)	Output Shape	Param #
dense_3 (Dense)	(None, 32)	416,128
dense_4 (Dense)	(None, 16)	528
dense_5 (Dense)	(None, 8)	136
dense_6 (Dense)	(None, 18)	162

Total params: 1,250,864 (4.77 MB)

Trainable params: 416,954 (1.59 MB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 833,910 (3.18 MB)

Training time: 119.02 seconds

Modified Model, with 4 hidden layers

Model: "sequential_2"

Layer (type)	Output Shape	Param #
dense_7 (Dense)	(None, 64)	832,256
dense_8 (Dense)	(None, 32)	2,080
dense_9 (Dense)	(None, 16)	528
dense_10 (Dense)	(None, 8)	136
dense_11 (Dense)	(None, 18)	162

Total params: 2,505,488 (9.56 MB)

Trainable params: 835,162 (3.19 MB)

Non-trainable params: 0 (0.00 B)

Optimizer params: 1,670,326 (6.37 MB)

Training time: 141.33 seconds