Task 1 Credit Card Fraud Detection

	Network Architecture	Training Parameters	Training Accuracy	Test Accuracy
1	(50,)	activation: relu, learning_rate: constant, learning_rate_init: 0.001, batch_size: auto	0.999648883449369	0.9995389580451821
2	(100,)	activation: relu, learning_rate: adaptive, learning_rate_init: 0.01, batch_size: 32	0.9994131337653739	0.9995034932794269
3	(50, 50)	activation: tanh, learning_rate: constant, learning_rate_init: 0.001, batch_size: 64	0.9997893300696215	0.9993970989821612
4	(100, 50)	activation: tanh, learning_rate: adaptive, learning_rate_init: 0.01, batch_size: 64	0.9994031018639273	0.9995034932794269
5	(50, 50, 50)	activation: logistic, learning_rate: constant, learning_rate_init: 0.001, batch_size: 32	0.9994883730262234	0.9995744228109373
6	(100, 100)	activation: relu, learning_rate: constant, learning_rate_init: 0.01, batch_size: 64	0.9993930699624807	0.9995389580451821
7	(50, 100, 50)	activation: tanh, learning_rate: adaptive, learning_rate_init: 0.005, batch_size: 32	0.9993880540117573	0.9993970989821612
8	(100, 50, 25)	activation: relu, learning_rate: adaptive, learning_rate_init: 0.001, batch_size: 64	0.9997642503160049	0.9996453523424478
9	(200,)	activation: logistic, learning_rate: constant, learning_rate_init: 0.01, batch_size: auto	0.9997040589073253	0.9994680285136717

| 10 | (100, 100, 100) | activation: relu, learning_rate: adaptive, learning_rate_init: 0.001, batch_size: 32 | 0.9997592343652816 | 0.9996098875766926 |

Key Findings:

The experiment highlighted that the configuration with (100, 50, 25) hidden layers, relu activation, adaptive learning rate, and learning_rate_init of 0.001 achieved the highest training accuracy (0.999764) and test accuracy (0.999645). This setup demonstrates exceptional performance in distinguishing fraudulent transactions, outperforming other configurations. The results indicate that a more complex network architecture with a well-chosen activation function and adaptive learning rate can lead to better generalization. However, all tested models exhibited very high accuracy, suggesting that while complexity can enhance performance, simpler models also performed well.

	Network Architecture	Training Parameters	Features Used	Training MSE	Test MSE
1	(50,)	Activation: relu, Learning Rate: constant, Learning Rate Init: 0.001, Batch Size: auto	Sex, Length, Diameter, Height, Whole weight, Shucked weight	4.3669	4.8895
2	(100,)	Activation: relu, Learning Rate: adaptive, Learning Rate Init: 0.01, Batch Size: 32	Sex, Length, Diameter, Height	4.9253	5.2501
3	(50, 50)	Activation: tanh, Learning Rate: constant, Learning Rate Init: 0.001, Batch Size: 64	Sex, Whole weight, Shucked weight, Viscera weight, Shell weight	4.0406	4.6123
4	(100, 50)	Activation: tanh, Learning Rate: adaptive, Learning Rate Init: 0.01, Batch Size: 64	Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, Shell weight	3.7795	4.6736
5	(50, 50, 50)	Activation: logistic, Learning Rate: constant, Learning Rate Init: 0.001, Batch Size: 32	Sex, Length, Diameter, Height, Whole weight, Shucked weight	4.2702	4.8244

6	(100, 100)	Activation: relu, Learning Rate: constant, Learning Rate Init: 0.01, Batch Size: 64	Sex, Length, Diameter, Height	4.0766	4.7331
7	(50, 100, 50)	Activation: tanh, Learning Rate: adaptive, Learning Rate Init: 0.005, Batch Size: 32	Sex, Whole weight, Shucked weight, Viscera weight, Shell weight	1.3972	7.1212
8	(100, 50, 25)	Activation: relu, Learning Rate: adaptive, Learning Rate Init: 0.001, Batch Size: 64	Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, Shell weight	3.8660	4.6811
9	(200,)	Activation: logistic, Learning Rate: constant, Learning Rate Init: 0.01, Batch Size: auto	Sex, Length, Diameter, Height, Whole weight, Shucked weight	4.3956	4.9594
10	(100, 100, 100)	Activation: relu, Learning Rate: adaptive, Learning Rate Init: 0.001, Batch Size: 32	Sex, Length, Diameter, Height	2.1129	6.9743
11	(100, 75)	Activation: relu, Learning Rate: adaptive, Learning Rate Init: 0.005, Batch Size: 64	Sex, Whole weight, Shucked weight, Viscera weight, Shell weight	4.6611	5.2659
12	(150, 100, 50)	Activation: relu, Learning Rate: constant, Learning Rate Init: 0.01, Batch Size: 32	Length, Diameter, Height, Whole weight, Shucked weight, Viscera weight, Shell weight	4.3821	5.1303
13	(200, 100)	Activation: tanh, Learning Rate: adaptive, Learning Rate Init: 0.001, Batch Size: 32	Sex, Length, Diameter, Height	3.4049	4.8496
14	(50, 75, 50)	Activation: logistic, Learning Rate: adaptive, Learning	Sex, Length, Diameter, Height, Whole weight,	4.2030	4.8290

		Rate Init: 0.005, Batch Size: 64	Shucked weight		
15	(100, 100, 50, 25)	Activation: relu, Learning Rate: constant, Learning Rate Init: 0.001, Batch Size: 64	Sex, Whole weight, Shucked weight, Viscera weight, Shell weight	3.3442	5.0428

Key Findings:

The experiment revealed that the configuration with hidden layers (50, 100, 50), using the tanh activation function and adaptive learning rate, achieved the best training MSE of 1.397. However, its test MSE was notably higher at 7.121, indicating potential overfitting. The model with hidden layers (200, 100) using tanh and an adaptive learning rate achieved a lower test MSE of 4.849, but its training MSE was 3.404. This suggests a better balance but still leaves room for improvement. Fine-tuning parameters such as the learning rate, batch size, or experimenting with regularization techniques could help reduce the gap between training and test MSE, as well as reducing the overall MSE, potentially leading to more accurate and generalizable models.