113066542 陳柏安 HW3

Model training, Data Augmentation, Hyper-parameter Tuning

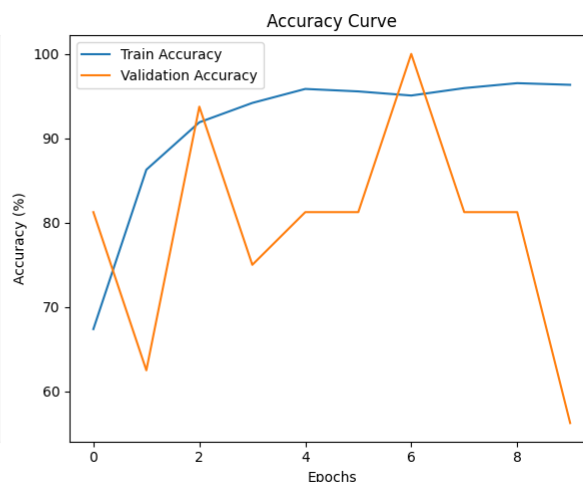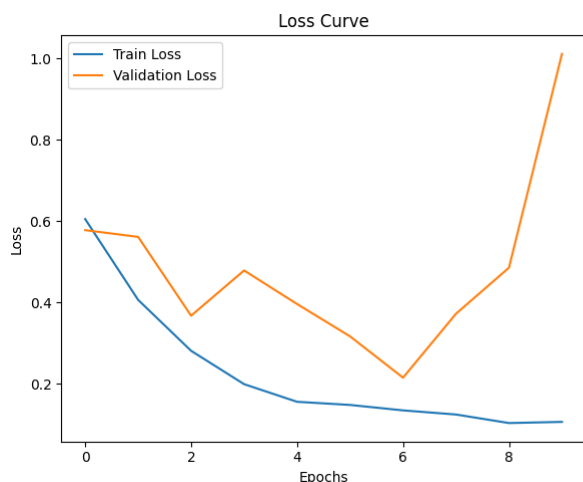**I. 測試紀錄**

Test 1：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=64, bias=True)
  (2): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.5, inplace=False)
  (5): Linear(in_features=64, out_features=64, bias=True)
  (6): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.5, inplace=False)
  (9): Linear(in_features=64, out_features=64, bias=True)
  (10): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.5, inplace=False)
  (13): Linear(in_features=64, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.6047 - train_acc: 67.38% - val_loss: 0.5774 - val_acc: 81.25% - time: 478.03s
Epoch 2/10 - loss: 0.4059 - train_acc: 86.28% - val_loss: 0.5609 - val_acc: 62.50% - time: 60.94s
Epoch 3/10 - loss: 0.2806 - train_acc: 91.89% - val_loss: 0.3670 - val_acc: 93.75% - time: 61.36s
Epoch 4/10 - loss: 0.1986 - train_acc: 94.19% - val_loss: 0.4783 - val_acc: 75.00% - time: 60.74s
Epoch 5/10 - loss: 0.1552 - train_acc: 95.85% - val_loss: 0.3957 - val_acc: 81.25% - time: 61.99s
Epoch 6/10 - loss: 0.1475 - train_acc: 95.56% - val_loss: 0.3162 - val_acc: 81.25% - time: 60.43s
Epoch 7/10 - loss: 0.1341 - train_acc: 95.07% - val_loss: 0.2146 - val_acc: 100.00% - time: 62.01s
Epoch 8/10 - loss: 0.1240 - train_acc: 95.95% - val_loss: 0.3718 - val_acc: 81.25% - time: 60.59s
Epoch 9/10 - loss: 0.1030 - train_acc: 96.53% - val_loss: 0.4851 - val_acc: 81.25% - time: 62.12s
Epoch 10/10 - loss: 0.1058 - train_acc: 96.34% - val_loss: 1.0109 - val_acc: 56.25% - time: 60.69s
```



```
Test Accuracy: 78.65%
Test Loss: 0.4748
```

在全部參數維持預設值情況下，進行第一次執行，從摺線圖分析，train loss 穩定下降，
train accuracy 穩定上升，但 validation loss 與 validation accuracy 後期表現過糟，波動情形
偏大，需再改進。

Test 2：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=64, bias=True)
  (2): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.3, inplace=False)
  (5): Linear(in_features=64, out_features=64, bias=True)
  (6): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.3, inplace=False)
  (9): Linear(in_features=64, out_features=64, bias=True)
  (10): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=64, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.4365 - train_acc: 85.01% - val_loss: 0.4873 - val_acc: 87.50% - time: 1539.08s
Epoch 2/10 - loss: 0.2392 - train_acc: 92.87% - val_loss: 0.5822 - val_acc: 62.50% - time: 62.76s
Epoch 3/10 - loss: 0.1731 - train_acc: 94.78% - val_loss: 0.3916 - val_acc: 81.25% - time: 62.16s
Epoch 4/10 - loss: 0.1203 - train_acc: 96.73% - val_loss: 0.3167 - val_acc: 93.75% - time: 62.10s
Epoch 5/10 - loss: 0.0915 - train_acc: 97.02% - val_loss: 0.8803 - val_acc: 50.00% - time: 63.37s
Epoch 6/10 - loss: 0.1031 - train_acc: 95.90% - val_loss: 0.7595 - val_acc: 56.25% - time: 61.48s
Epoch 7/10 - loss: 0.0948 - train_acc: 96.92% - val_loss: 0.3187 - val_acc: 87.50% - time: 62.66s
Epoch 8/10 - loss: 0.0769 - train_acc: 97.41% - val_loss: 1.8979 - val_acc: 50.00% - time: 61.62s
Epoch 9/10 - loss: 0.1054 - train_acc: 97.27% - val_loss: 0.7353 - val_acc: 62.50% - time: 62.87s
Epoch 10/10 - loss: 0.0793 - train_acc: 97.27% - val_loss: 1.7960 - val_acc: 56.25% - time: 61.63s
```
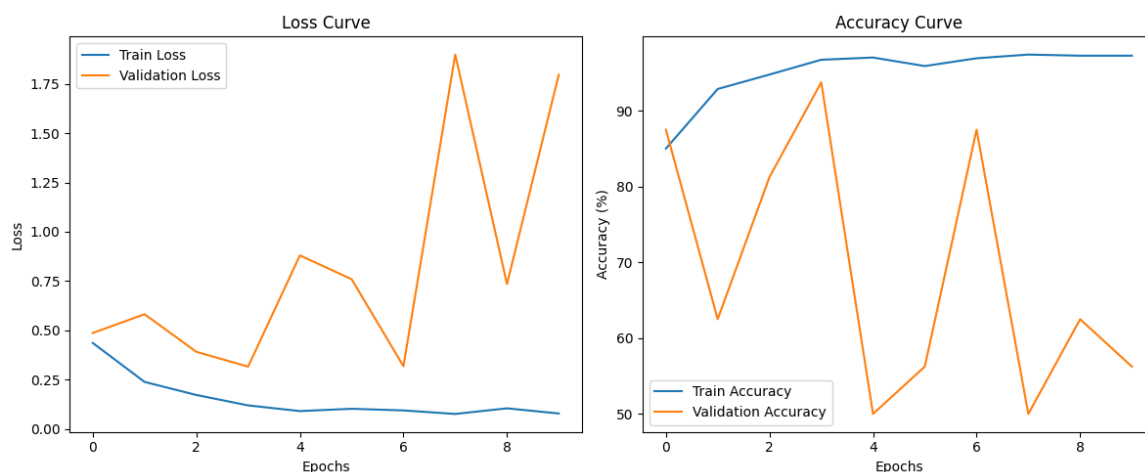


```
Test Accuracy: 75.05%
Test Loss: 0.5577
```

調整 dropout 全為 0.3 後，train loss & accuracy 皆有更加良好表現，但 validation loss & accuracy 在後期皆有較劇烈波動，應為過擬合的表現，因此在 test 的結果也略微降低。

Test 3：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=64, bias=True)
  (2): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=64, out_features=64, bias=True)
  (6): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.6, inplace=False)
  (9): Linear(in_features=64, out_features=64, bias=True)
  (10): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.6, inplace=False)
  (13): Linear(in_features=64, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.6278 - train_acc: 64.99% - val_loss: 0.5657 - val_acc: 81.25% - time: 980.66s
Epoch 2/10 - loss: 0.4679 - train_acc: 81.25% - val_loss: 0.4887 - val_acc: 81.25% - time: 63.03s
Epoch 3/10 - loss: 0.3401 - train_acc: 89.50% - val_loss: 0.5123 - val_acc: 75.00% - time: 61.96s
Epoch 4/10 - loss: 0.2460 - train_acc: 93.16% - val_loss: 0.4920 - val_acc: 75.00% - time: 62.94s
Epoch 5/10 - loss: 0.2105 - train_acc: 93.60% - val_loss: 0.3707 - val_acc: 81.25% - time: 61.60s
Epoch 6/10 - loss: 0.1546 - train_acc: 95.70% - val_loss: 0.4270 - val_acc: 81.25% - time: 62.87s
Epoch 7/10 - loss: 0.1742 - train_acc: 94.04% - val_loss: 0.5112 - val_acc: 75.00% - time: 61.78s
Epoch 8/10 - loss: 0.1415 - train_acc: 96.29% - val_loss: 0.5121 - val_acc: 75.00% - time: 63.33s
Epoch 9/10 - loss: 0.1494 - train_acc: 95.26% - val_loss: 0.3224 - val_acc: 81.25% - time: 61.63s
Epoch 10/10 - loss: 0.1329 - train_acc: 96.44% - val_loss: 0.3206 - val_acc: 87.50% - time: 62.33s
```
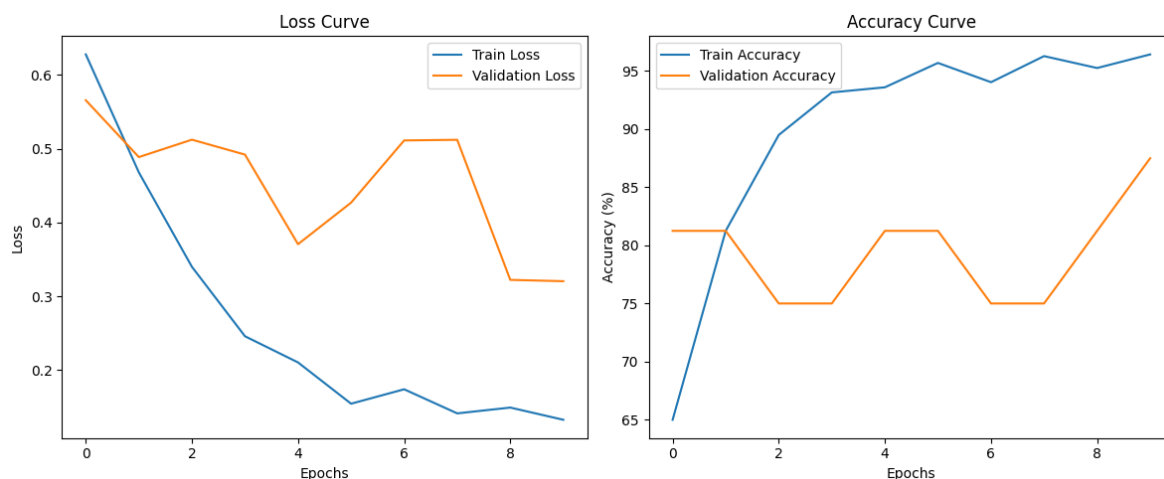


```
Test Accuracy: 82.08%
Test Loss: 0.4595
```

調整 dropout 全為 0.6 後，運算耗費時間大幅減少，train loss & accuracy 後期還是可以達到良好表現，又 validation loss & accuracy 在後期穩定性相較 **Test 1**(dropout 全為 0.5)好，因此最後得到的 test accuracy 略微提升。

Test 4：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=64, bias=True)
  (2): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=64, out_features=64, bias=True)
  (6): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=64, out_features=64, bias=True)
  (10): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=64, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.5256 - train_acc: 77.49% - val_loss: 0.5337 - val_acc: 75.00% - time: 64.06s
Epoch 2/10 - loss: 0.3306 - train_acc: 88.96% - val_loss: 0.4807 - val_acc: 81.25% - time: 61.91s
Epoch 3/10 - loss: 0.2267 - train_acc: 93.12% - val_loss: 0.3176 - val_acc: 87.50% - time: 63.28s
Epoch 4/10 - loss: 0.1843 - train_acc: 93.55% - val_loss: 0.4217 - val_acc: 87.50% - time: 61.91s
Epoch 5/10 - loss: 0.1494 - train_acc: 95.51% - val_loss: 0.6171 - val_acc: 68.75% - time: 63.12s
Epoch 6/10 - loss: 0.1488 - train_acc: 95.17% - val_loss: 1.2486 - val_acc: 50.00% - time: 61.78s
Epoch 7/10 - loss: 0.1264 - train_acc: 95.61% - val_loss: 0.7309 - val_acc: 50.00% - time: 63.17s
Epoch 8/10 - loss: 0.1323 - train_acc: 95.61% - val_loss: 0.3651 - val_acc: 81.25% - time: 61.64s
Epoch 9/10 - loss: 0.1184 - train_acc: 95.56% - val_loss: 0.4366 - val_acc: 75.00% - time: 62.82s
Epoch 10/10 - loss: 0.1086 - train_acc: 96.53% - val_loss: 0.1169 - val_acc: 100.00% - time: 61.46s
```
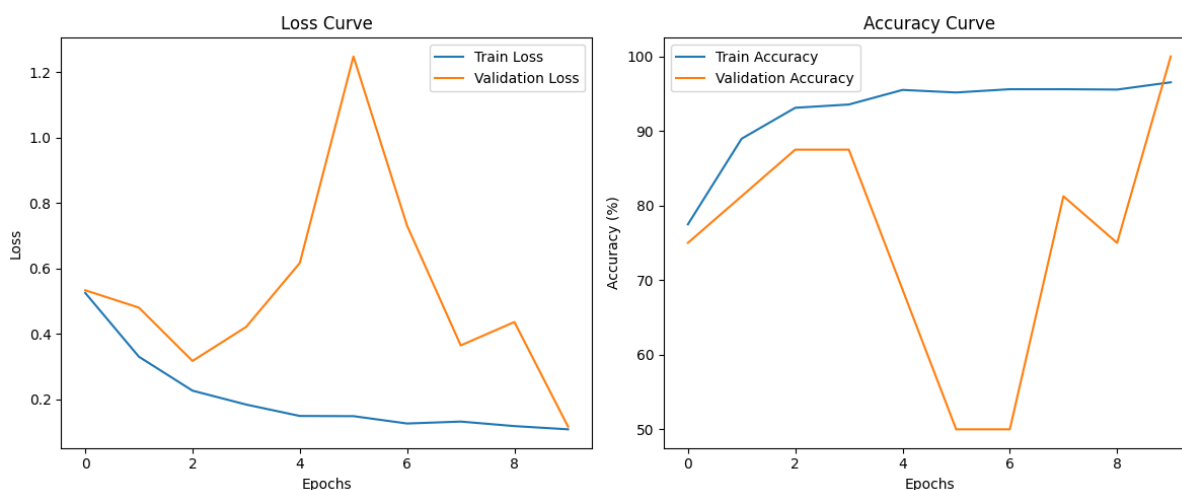


```
Test Accuracy: 83.18%
Test Loss: 0.4209
```

調整 dropout 依序分別為 0.6、0.4、0.3，train loss & accuracy 表現穩定，然而 validation loss & accuracy 在某些 epoch 波動過大，穩定性不佳，最後得到的 test accuracy 相較前一次 (**Test 3**)些微提升。

Test 5：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=32, bias=True)
  (2): BatchNorm1d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=32, out_features=32, bias=True)
  (6): BatchNorm1d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=32, out_features=32, bias=True)
  (10): BatchNorm1d(32, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=32, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.5950 - train_acc: 70.75% - val_loss: 0.6595 - val_acc: 56.25% - time: 63.71s
Epoch 2/10 - loss: 0.4811 - train_acc: 81.98% - val_loss: 0.6179 - val_acc: 68.75% - time: 63.69s
Epoch 3/10 - loss: 0.3775 - train_acc: 87.60% - val_loss: 0.6080 - val_acc: 68.75% - time: 61.73s
Epoch 4/10 - loss: 0.3015 - train_acc: 90.43% - val_loss: 0.5454 - val_acc: 68.75% - time: 63.06s
Epoch 5/10 - loss: 0.2455 - train_acc: 92.77% - val_loss: 0.5023 - val_acc: 81.25% - time: 61.54s
Epoch 6/10 - loss: 0.2192 - train_acc: 93.46% - val_loss: 0.6472 - val_acc: 50.00% - time: 62.76s
Epoch 7/10 - loss: 0.1975 - train_acc: 93.70% - val_loss: 0.4930 - val_acc: 75.00% - time: 61.77s
Epoch 8/10 - loss: 0.1710 - train_acc: 94.63% - val_loss: 0.7890 - val_acc: 56.25% - time: 62.85s
Epoch 9/10 - loss: 0.1583 - train_acc: 94.97% - val_loss: 0.5180 - val_acc: 75.00% - time: 61.99s
Epoch 10/10 - loss: 0.1767 - train_acc: 94.34% - val_loss: 0.7185 - val_acc: 62.50% - time: 62.76s
```
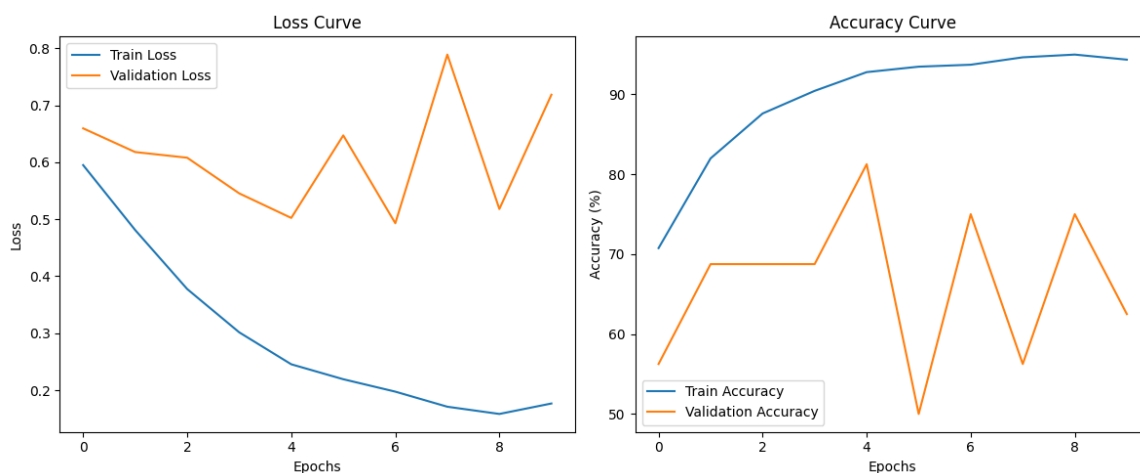


```
Test Accuracy: 76.77%
Test Loss: 0.4495
```

維持 dropout 分別依序為 0.6、0.4、0.3，同時一併減少每層計算的神經元數，從 64 降為 32，train loss & accuracy 表現漸進趨於穩定，但 validation loss & accuracy 明顯偏差過大，應為欠擬合的情形，最後得到的 test accuracy 也不夠理想。

Test 6：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=128, bias=True)
  (2): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=128, out_features=128, bias=True)
  (6): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=128, out_features=128, bias=True)
  (10): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=128, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
Epoch 1/10 - loss: 0.4449 - train_acc: 81.64% - val_loss: 0.4244 - val_acc: 87.50% - time: 63.14s
Epoch 2/10 - loss: 0.2313 - train_acc: 91.94% - val_loss: 0.5233 - val_acc: 62.50% - time: 63.32s
Epoch 3/10 - loss: 0.1541 - train_acc: 94.68% - val_loss: 0.3231 - val_acc: 93.75% - time: 63.21s
Epoch 4/10 - loss: 0.1321 - train_acc: 95.17% - val_loss: 0.7067 - val_acc: 68.75% - time: 61.92s
Epoch 5/10 - loss: 0.1077 - train_acc: 95.75% - val_loss: 0.2370 - val_acc: 100.00% - time: 63.08s
Epoch 6/10 - loss: 0.0931 - train_acc: 96.34% - val_loss: 0.2784 - val_acc: 93.75% - time: 61.94s
Epoch 7/10 - loss: 0.0912 - train_acc: 96.39% - val_loss: 0.3572 - val_acc: 81.25% - time: 62.92s
Epoch 8/10 - loss: 0.1049 - train_acc: 96.09% - val_loss: 1.4453 - val_acc: 50.00% - time: 62.66s
Epoch 9/10 - loss: 0.1190 - train_acc: 95.36% - val_loss: 0.3389 - val_acc: 87.50% - time: 63.00s
Epoch 10/10 - loss: 0.1261 - train_acc: 94.97% - val_loss: 0.8615 - val_acc: 62.50% - time: 62.62s
```
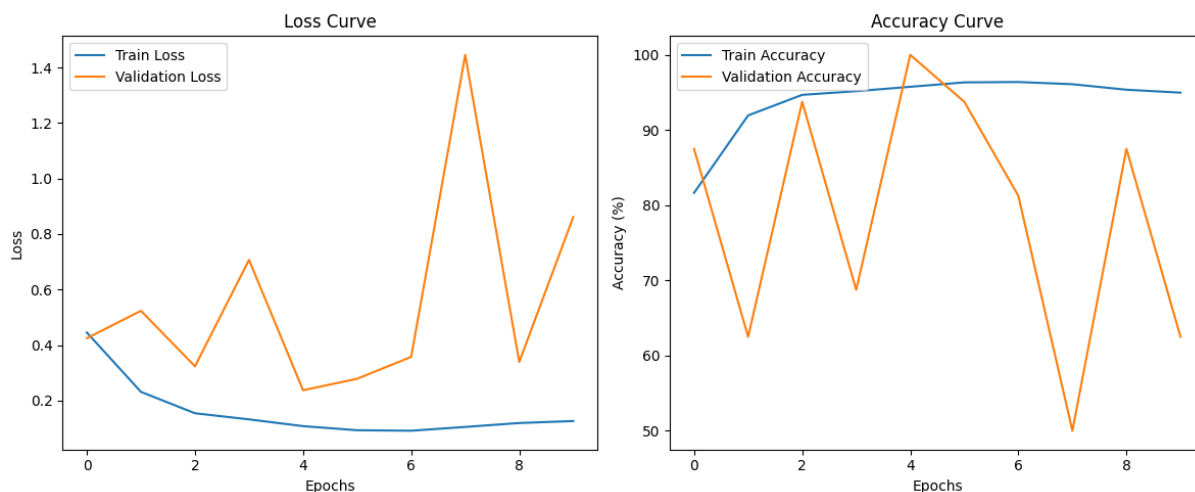


```
Test Accuracy: 81.93%
Test Loss: 0.4333
```

維持 dropout 分別依序為 0.6、0.4、0.3，改為一併增加每層計算的神經元數，從 64 增為 128，train loss & accuracy 表現穩定良好，但 validation loss & accuracy 波動情形過大，極為不穩定，應為過擬合的情形。

Test 7：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=128, bias=True)
  (2): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=128, out_features=128, bias=True)
  (6): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=128, out_features=128, bias=True)
  (10): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=128, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
# hyperparameter
lr = 0.0001
weight_decay = 0.001
epochs = 10
```

```
Epoch 1/10 - loss: 0.0956 - train_acc: 97.27% - val_loss: 0.1777 - val_acc: 100.00% - time: 63.70s
Epoch 2/10 - loss: 0.0891 - train_acc: 97.27% - val_loss: 0.3238 - val_acc: 81.25% - time: 61.64s
Epoch 3/10 - loss: 0.0747 - train_acc: 97.61% - val_loss: 0.2751 - val_acc: 81.25% - time: 63.62s
Epoch 4/10 - loss: 0.0716 - train_acc: 97.61% - val_loss: 0.2995 - val_acc: 87.50% - time: 62.20s
Epoch 5/10 - loss: 0.0810 - train_acc: 97.31% - val_loss: 0.1425 - val_acc: 100.00% - time: 63.86s
Epoch 6/10 - loss: 0.0597 - train_acc: 98.05% - val_loss: 0.2128 - val_acc: 87.50% - time: 63.52s
Epoch 7/10 - loss: 0.0586 - train_acc: 98.49% - val_loss: 0.7329 - val_acc: 68.75% - time: 62.33s
Epoch 8/10 - loss: 0.0527 - train_acc: 98.34% - val_loss: 0.0735 - val_acc: 100.00% - time: 63.04s
Epoch 9/10 - loss: 0.0602 - train_acc: 98.05% - val_loss: 0.0855 - val_acc: 100.00% - time: 61.44s
Epoch 10/10 - loss: 0.0665 - train_acc: 97.71% - val_loss: 0.0805 - val_acc: 100.00% - time: 63.62s
```
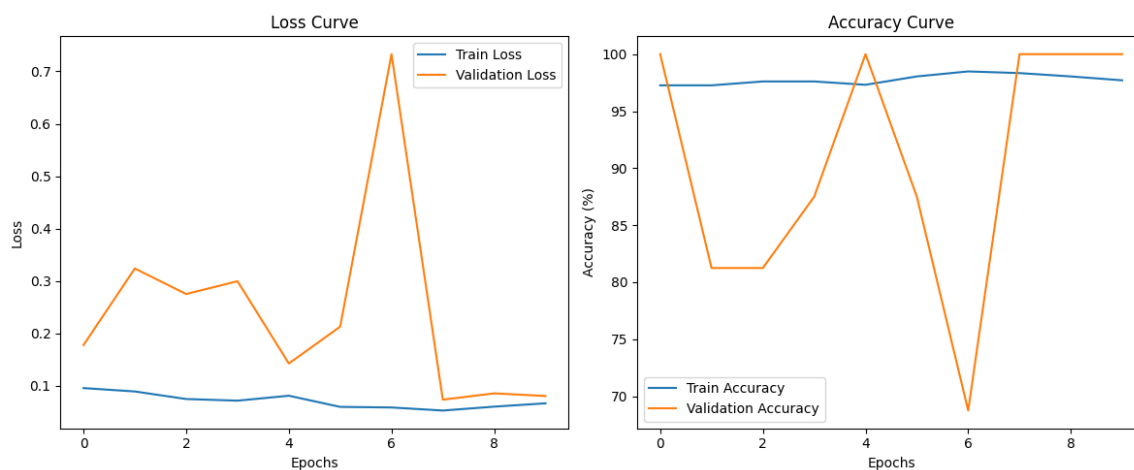


```
Test Accuracy: 83.85%
Test Loss: 0.4179
```

維持前一次(**Test 6**) model 設定，唯獨改變 learning rate 為 0.0001，train loss & accuracy 穩定度大幅提升，但 validation loss & accuracy 波動情形仍然過大，極為不穩定，應為過擬合的情形，需再改進，而得到結果 test accuracy 則為目前最佳。

Test 8：

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=128, bias=True)
  (2): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=128, out_features=128, bias=True)
  (6): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=128, out_features=128, bias=True)
  (10): BatchNorm1d(128, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=128, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
# hyperparameter
lr = 0.0001
weight_decay = 0.0001
epochs = 10
```

```
Epoch 1/10 - loss: 0.0973 - train_acc: 96.97% - val_loss: 0.3186 - val_acc: 81.25% - time: 62.11s
Epoch 2/10 - loss: 0.0760 - train_acc: 97.90% - val_loss: 0.3309 - val_acc: 75.00% - time: 63.22s
Epoch 3/10 - loss: 0.0719 - train_acc: 97.80% - val_loss: 0.1735 - val_acc: 100.00% - time: 61.64s
Epoch 4/10 - loss: 0.0794 - train_acc: 97.17% - val_loss: 0.1501 - val_acc: 93.75% - time: 63.65s
Epoch 5/10 - loss: 0.0629 - train_acc: 98.10% - val_loss: 0.1741 - val_acc: 87.50% - time: 62.01s
Epoch 6/10 - loss: 0.0603 - train_acc: 98.00% - val_loss: 0.1180 - val_acc: 100.00% - time: 63.59s
Epoch 7/10 - loss: 0.0581 - train_acc: 98.05% - val_loss: 0.3138 - val_acc: 87.50% - time: 63.15s
Epoch 8/10 - loss: 0.0633 - train_acc: 98.14% - val_loss: 0.2371 - val_acc: 87.50% - time: 62.71s
Epoch 9/10 - loss: 0.0474 - train_acc: 98.63% - val_loss: 0.6334 - val_acc: 75.00% - time: 63.27s
Epoch 10/10 - loss: 0.0487 - train_acc: 98.29% - val_loss: 0.1322 - val_acc: 93.75% - time: 61.54s
```
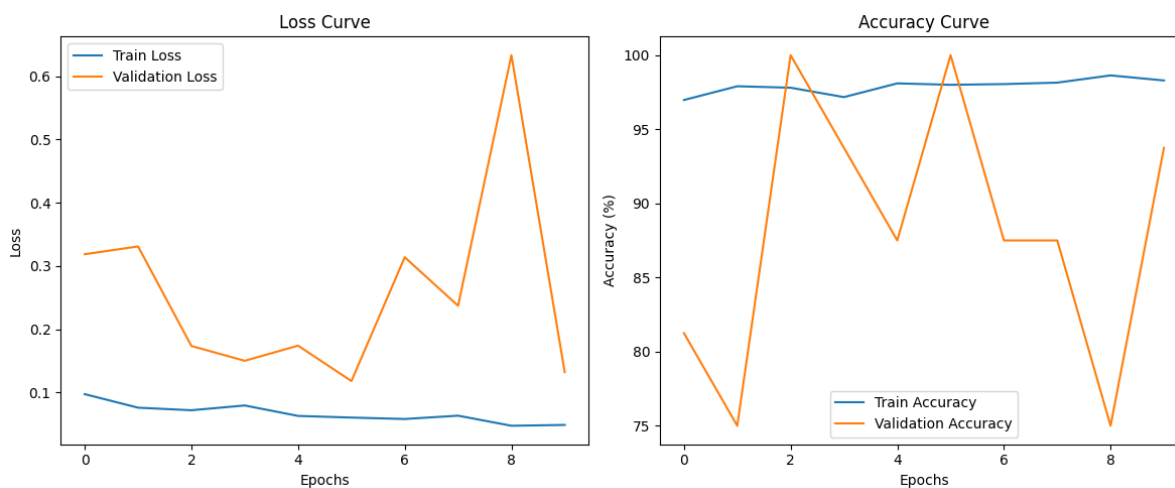

```
Test Accuracy: 83.23%
Test Loss: 0.4815
```

維持前一次(**Test 7**) model 設定，唯獨改變 weight decay 為 0.0001，然而 validation loss & accuracy 的數值表現並無明顯改善，仍需做其他參數調整改進。

**Test 9：**

```
Sequential(
  (0): Flatten(start_dim=1, end_dim=-1)
  (1): Linear(in_features=65536, out_features=64, bias=True)
  (2): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (3): ReLU()
  (4): Dropout(p=0.6, inplace=False)
  (5): Linear(in_features=64, out_features=64, bias=True)
  (6): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (7): ReLU()
  (8): Dropout(p=0.4, inplace=False)
  (9): Linear(in_features=64, out_features=64, bias=True)
  (10): BatchNorm1d(64, eps=1e-05, momentum=0.1, affine=True, track_running_stats=True)
  (11): ReLU()
  (12): Dropout(p=0.3, inplace=False)
  (13): Linear(in_features=64, out_features=1, bias=True)
  (14): Sigmoid()
)
```

```
# hyperparameter
lr = 0.0001
weight_decay = 0.0001
epochs = 10
```

```
Epoch 1/10 - loss: 0.6366 - train_acc: 65.09% - val_loss: 0.6307 - val_acc: 75.00% - time: 867.22s
Epoch 2/10 - loss: 0.5652 - train_acc: 75.05% - val_loss: 0.5925 - val_acc: 75.00% - time: 61.85s
Epoch 3/10 - loss: 0.4990 - train_acc: 83.84% - val_loss: 0.5695 - val_acc: 81.25% - time: 63.32s
Epoch 4/10 - loss: 0.4349 - train_acc: 89.75% - val_loss: 0.5494 - val_acc: 87.50% - time: 64.03s
Epoch 5/10 - loss: 0.3921 - train_acc: 92.14% - val_loss: 0.4433 - val_acc: 93.75% - time: 63.27s
Epoch 6/10 - loss: 0.3492 - train_acc: 94.53% - val_loss: 0.5132 - val_acc: 93.75% - time: 64.99s
Epoch 7/10 - loss: 0.3181 - train_acc: 94.53% - val_loss: 0.4570 - val_acc: 81.25% - time: 64.09s
Epoch 8/10 - loss: 0.2876 - train_acc: 95.41% - val_loss: 0.3995 - val_acc: 100.00% - time: 63.00s
Epoch 9/10 - loss: 0.2617 - train_acc: 96.44% - val_loss: 0.4228 - val_acc: 93.75% - time: 64.23s
Epoch 10/10 - loss: 0.2423 - train_acc: 96.58% - val_loss: 0.4469 - val_acc: 81.25% - time: 63.08s
```
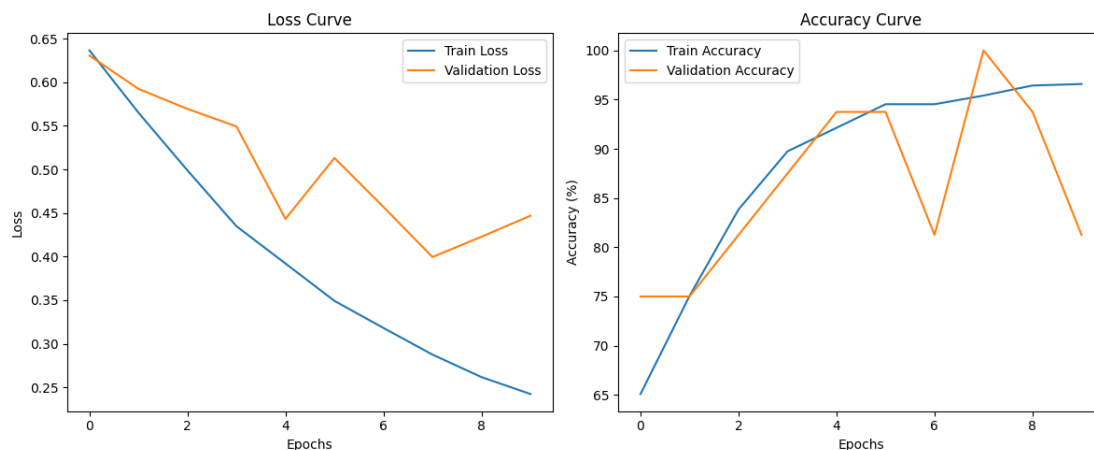


```
Test Accuracy: 81.61%
Test Loss: 0.4304
```

weight decay 與 learning rate 皆維持 0.0001，將每層計算的神經元數從 128 降回 64，training set 表現略降些，但 validation loss 與 validation accuracy 劇烈波動情也有些許改善。目前模型在驗證集上的表現波動還是大，代表模型仍存在過擬合的情況，應再嘗試調整 hyperparameter 設定。