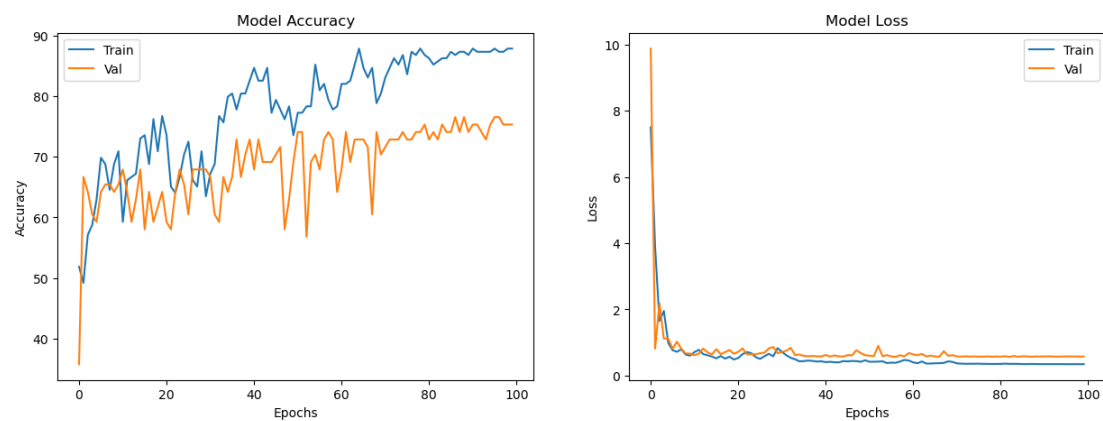


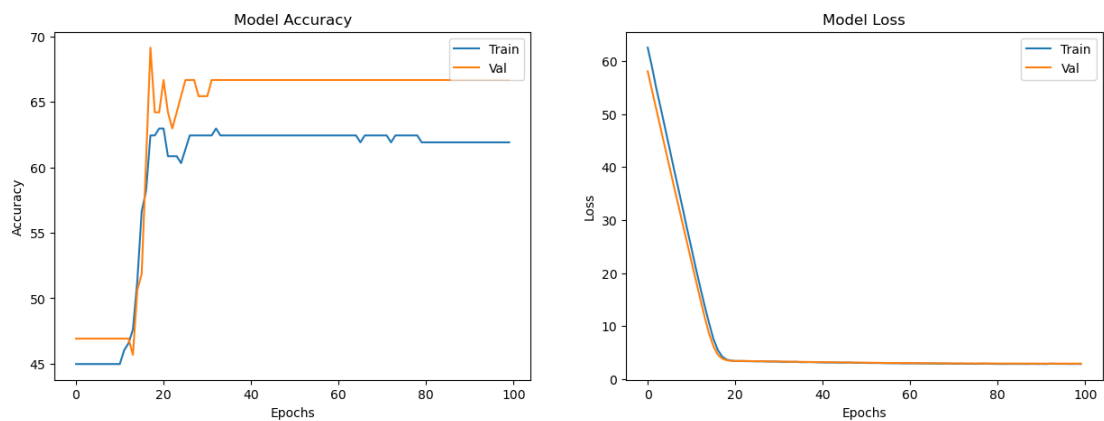
1. 選擇的超參數：神經網路層數（1、5、10）及訓練迴圈數（50、150、200）

	Training accuracy	Validation accuracy	Best Validation accuracy	Test accuracy
Original (# layer = 3, epochs = 100)	80.42 %	72.84 %	76.54 %	67.74 %
# layer = 1	62.96 %	66.67 %	66.67 %	64.52 %
# layer = 5	85.19 %	71.60 %	72.84 %	74.19 %
# layer = 10	80.95 %	67.90 %	79.01 %	61.29 %
epochs = 50	80.42 %	70.37 %	70.37 %	58.06 %
epochs = 150	87.83 %	77.78 %	80.25 %	80.65 %
epochs = 200	88.36 %	81.48 %	81.48 %	70.97 %

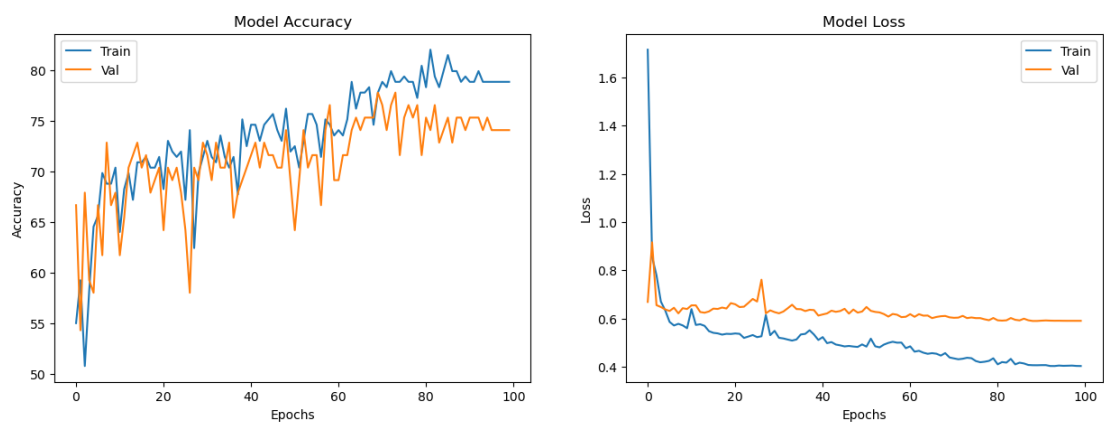
表格一



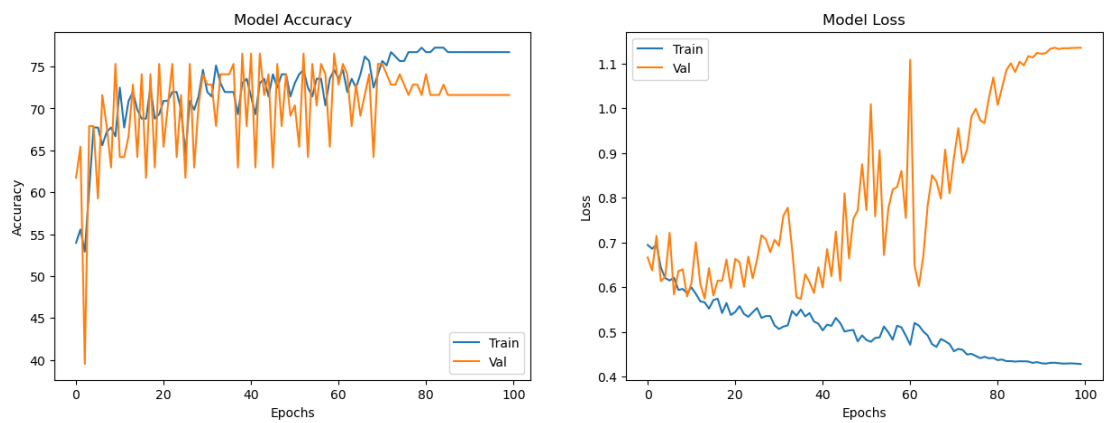
圖一、Original



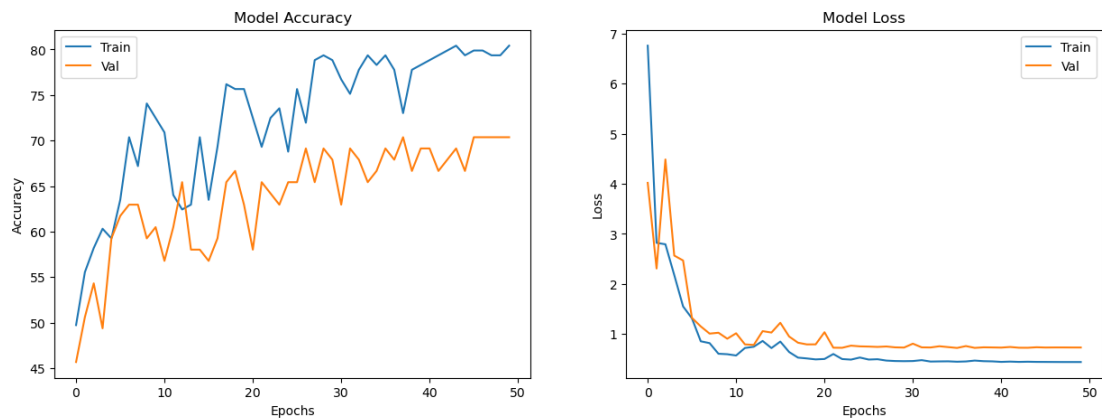
圖二、# layer = 1



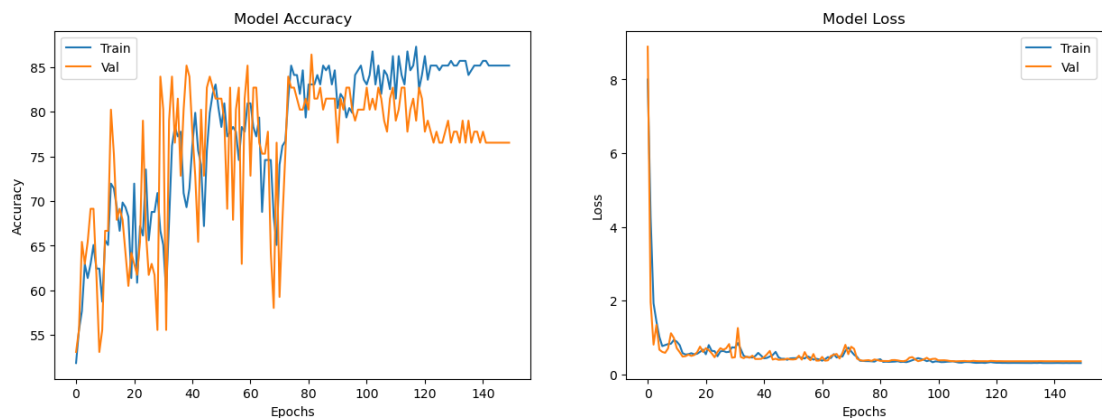
圖三、# layer = 5



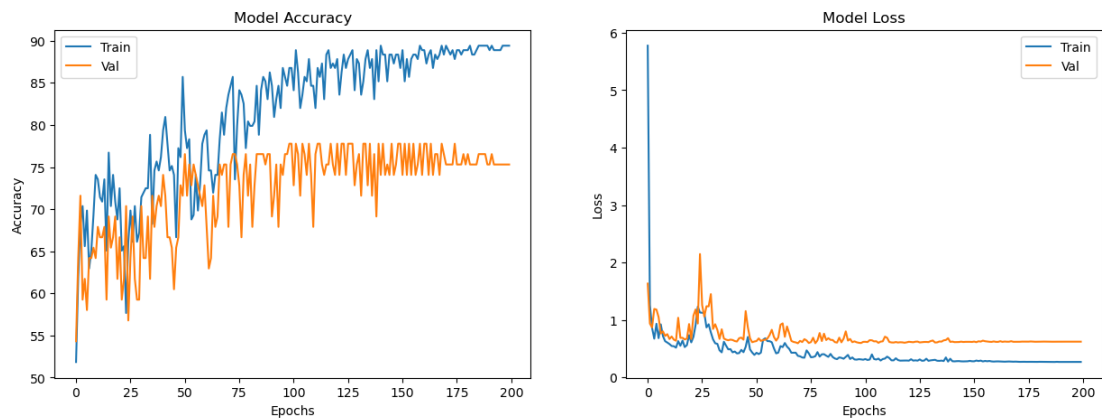
圖四、# layer = 10



圖五、epochs = 50



圖六、epochs = 150



圖七、epochs = 200

- 從表格一與圖二三四可以看出，只有 1 層時準確度都是下降；變為 5 層時，訓練及測試準確度上升了 5~6%；但增加至 10 層時驗證及測試準確度又減少了。由此可知，增加層數有機會讓模型表現更好，但增加過多可能反而使表現變差。從圖五六七發現增加 epochs 可以讓準確度上升，但並非 epochs 越多準確度越高。

3. 從表格二中可以看出，大部分實驗的驗證準確度皆低於訓練準確度，表示模型發生過擬合，變得過度依賴訓練集的資料，因此在測試集上表現就沒那麼好。而再看到表格三發現驗證集與測試集的準確度也有些許落差，可能意味著測試集和驗證集中的資料分布不盡相同。

	Training accuracy	Validation accuracy	Difference
Original (# layer = 3, epochs = 100)	80.42 %	72.84 %	7.58 %
# layer = 1	62.96 %	66.67 %	-3.71 %
# layer = 5	85.19 %	71.60 %	13.59 %
# layer = 10	80.95 %	67.90 %	13.05 %
epochs = 50	80.42 %	70.37 %	10.05 %
epochs = 150	87.83 %	77.78 %	10.05 %
epochs = 200	88.36 %	81.48 %	6.88 %

表格二

	Validation accuracy	Test accuracy	Difference
Original (# layer = 3, epochs = 100)	72.84 %	67.74 %	-5.10 %
# layer = 1	66.67 %	64.52 %	-2.15 %
# layer = 5	71.60 %	74.19 %	2.59 %
# layer = 10	67.90 %	61.29 %	-6.61 %
epochs = 50	70.37 %	58.06 %	-12.31 %
epochs = 150	77.78 %	80.65 %	2.87 %
epochs = 200	81.48 %	70.97 %	-10.51 %

表格三

4. 進行特徵篩選有幾個常用的方法，像是過濾法（利用皮爾森相關係數或卡方檢定等，選出前 k 個值最高的特徵）、包裝器特徵篩選方法（例如隨機森林、kNN）、嵌入法（例如 L1, L2 正規化），目的是為了減少對預測沒有幫助的特徵避免模型過擬合，同時能夠增加可解釋性。

參考資料：<https://zhuanlan.zhihu.com/p/74198735>、

<https://medium.com/kkproject/%E7%89%B9%E5%BE%B5%E7%AF%A9%E9%81%B8-%E5%9F%BA%E6%9C%AC%E4%BB%8B%E7%B4%B9-5645dc8cae56>

5. 在深度學習的模型中有一些是專門用來處理表格資料，如 TabNet, Wide & Deep Learning, DeepFM 等。以 TabNet 為例，它先是建構一個可以模擬決策樹的神經網路，保留樹模型的優勢（如決策流形為超平面邊界、可解釋性高），再來引進了注意力機制讓模型自己學習特徵之間的關聯性，進而節省在特徵工程上所耗費的時間與人力，並在訓練過程中利用遞歸特徵消除法逐步刪除較不重要的特徵，從而提高模型的性能。

參考資料：<https://zhuanlan.zhihu.com/p/152211918>、

<https://zhuanlan.zhihu.com/p/447814132>