HW2

郭又瑜 113034547

1. (20 pts) Select 2 hyper-parameters of the artificial neural network used in Lab 2 and set 3 different values for each. Perform experiments to compare the effects of varying these hyper-parameters on the loss and accuracy metrics across the training, validation, and test datasets. Present your findings with appropriate tables.

batch_size = 16:

	Train loss	Train acc	Val loss	Val acc
lr (learning rate) = 0.001	0.3537	85.1852%	0.3326	85.1852%
lr = 0.0001	0.5022	74.6032%	0.4392	74.0741%
lr = 0.005	0.3485	84.6561%	0.2937	87.6543%

batch_size = 32:

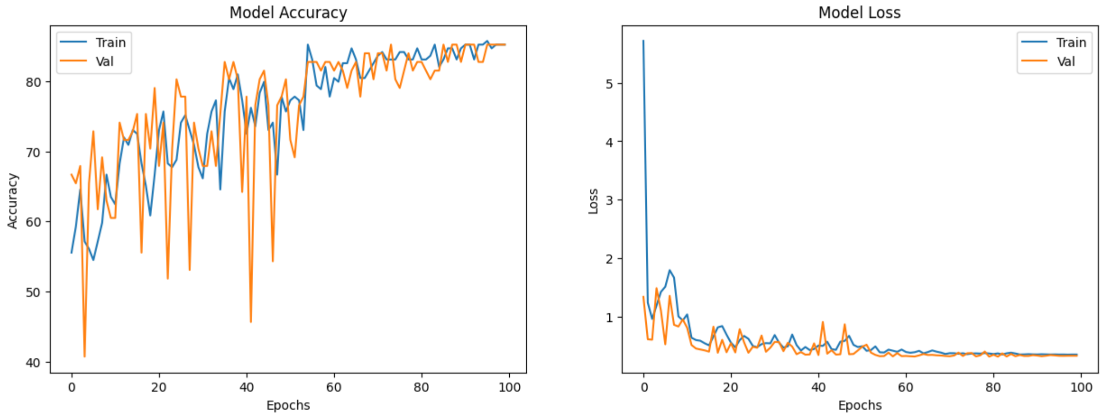
	Train loss	Train acc	Val loss	Val acc
lr (learning rate) = 0.001	0.4113	80.9524%	0.4434	82.7160%
lr = 0.0001	0.5021	76.1905%	0.4892	74.0741%
lr = 0.005	0.3240	85.7143%	0.3665	86.4198%

batch_size = 64:

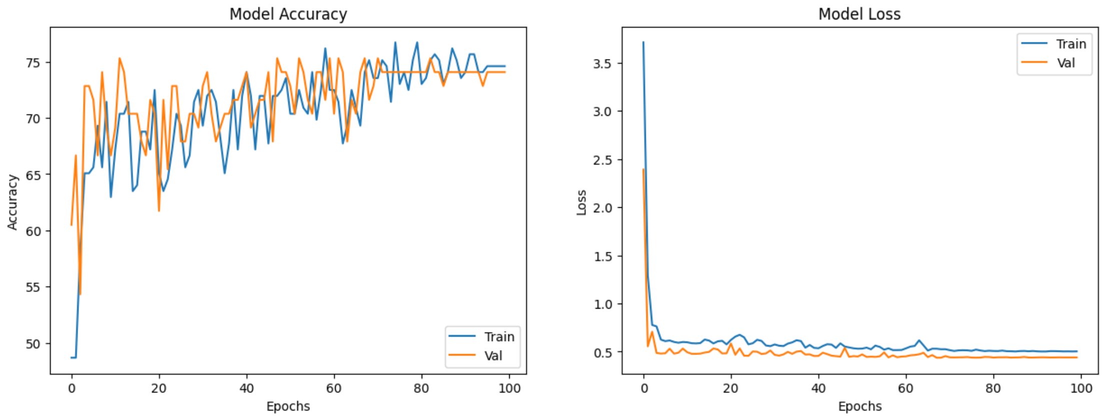
	Train loss	Train acc	Val loss	Val acc
lr (learning rate) = 0.001	0.4885	76.7196%	0.4812	77.7778%
lr = 0.0001	0.5522	71.9577%	0.5353	70.3704%
lr = 0.005	0.4471	79.8942%	0.4338	80.2469%

2. (20 pts) Based on your experiments in Question 1, analyze the outcomes. What differences do you observe with the changes in hyper-parameters? Discuss whether these adjustments contributed to improvements in model performance, you can use plots to support your points. (Approximately 100 words.)

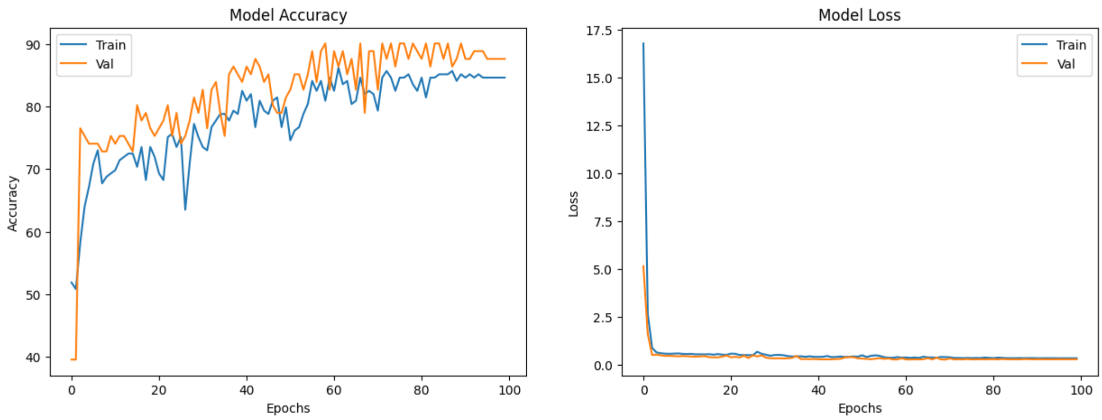
batch_size = 16, lr = 0.001



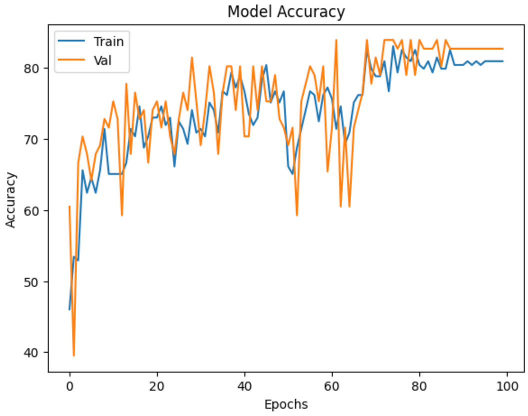
batch_size = 16, lr = 0.0001

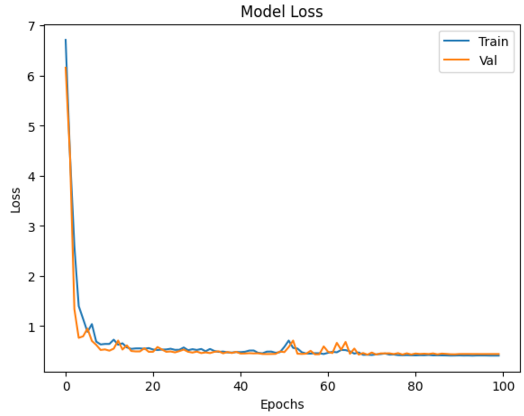


batch_size = 16, lr = 0.005

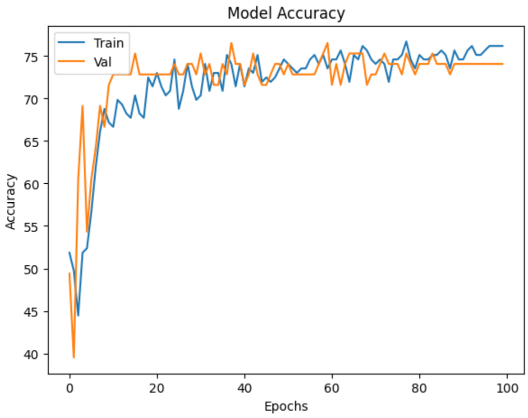


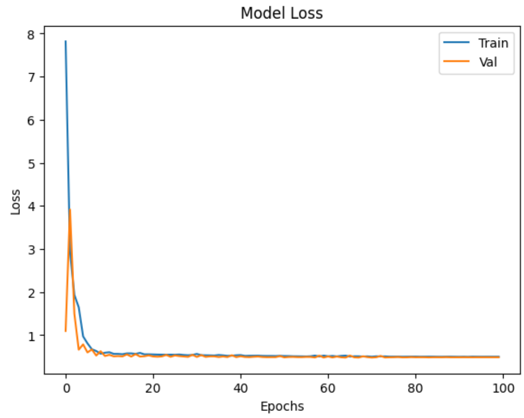
batch_size = 32, lr = 0.001



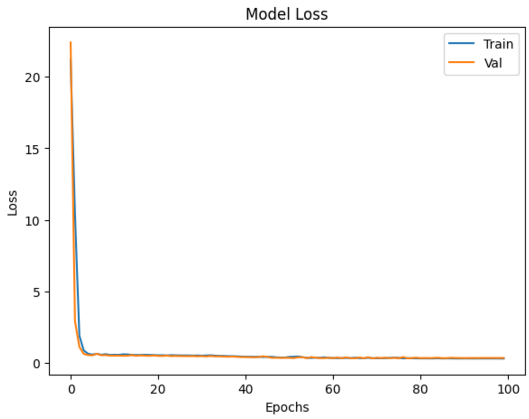


batch_size = 32, lr = 0.0001

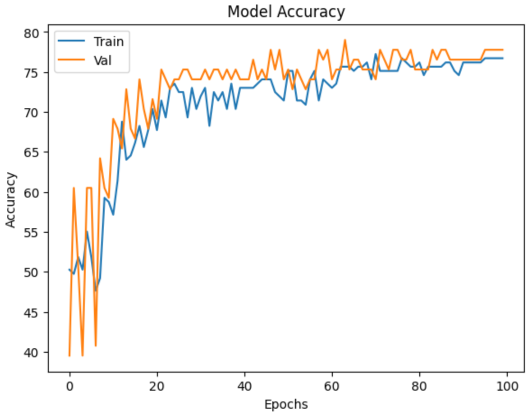


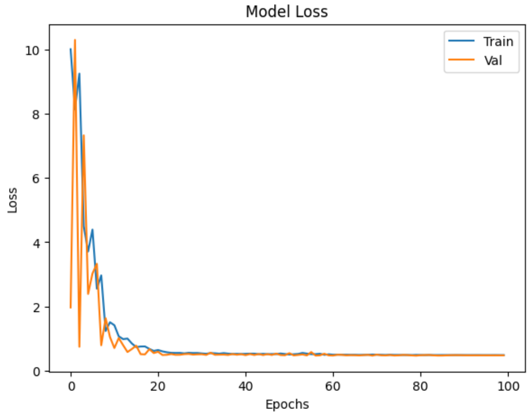


batch_size = 32, lr = 0.005

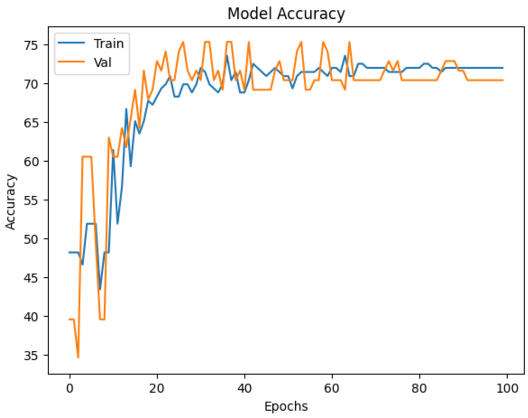


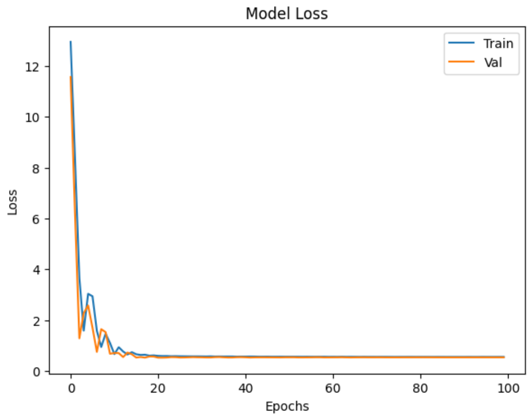
batch_size = 64, lr = 0.001



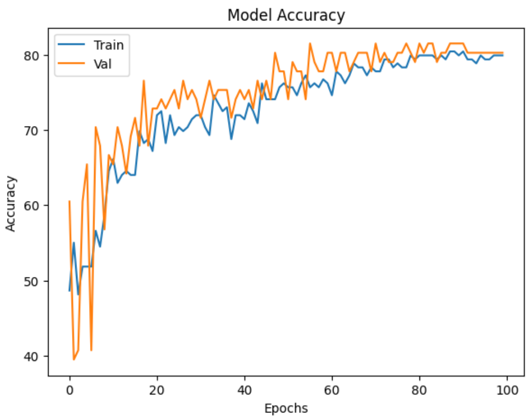


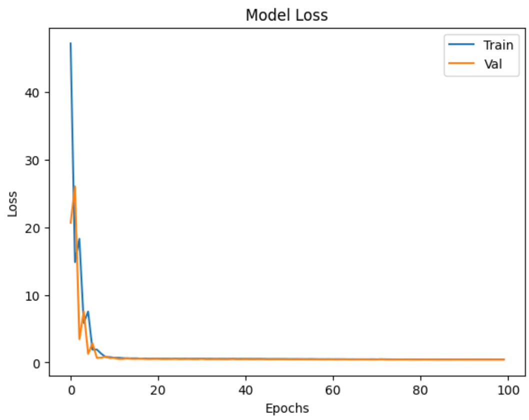
batch_size = 64, lr = 0.0001





batch_size = 64, lr = 0.005





學習率 (lr) 的影響：

- lr = 0.0001：最低的學習率，所有批次大小的驗證準確率皆偏低。由圖可知，模型學習速度極慢，準確率上升緩慢，顯示模型無法有效學習特徵。且驗證準確率波動大，顯示模型的泛化能力較差。
- lr = 0.001：中等學習率，此設定下，訓練與驗證損失下降平穩，模型學習穩定。

- lr = 0.005：在所有批次大小下表現最佳，驗證準確率最高可達 87.65% (batch size = 16)。

批次大小 (batch_size) 的影響

- 小批次具有較高的驗證準確率，顯示較小批次的梯度更新較頻繁，有助於模型泛化，但是準確率波動較大。
- 較大的批次則普遍表現較差，推測是因為較少的梯度更新次數影響了模型的泛化能力。
- batch_size = 16 搭配 lr = 0.005 時達到最佳準確率（87.65%），顯示小批次訓練在較高學習率下效果更佳。
- 大批次雖然學習過程較平穩，但準確率較低，權重更新次數較少導致學習效果受限。

3. (20 pts) In Lab 2, you may have noticed a discrepancy in accuracy between the training and test datasets. What do you think causes this occurrence? Discuss potential reasons for the gap in accuracy. (Approximately 100 words.)

訓練集與測試集的準確率存在差異通常與以下幾個因素有關：

- 資料分佈不同：

  測試集的特徵分佈與訓練集不同，例如某些類別的比例不同或資料收集方式不同，模型的泛化能力會下降，導致測試準確率低。

- 資料集大小：

  若訓練資訊量不足，模型可能無法學習到足夠的泛化特徵，導致在測試集上準確率較低。

- 測試集的多樣性不足：

  如果測試集資訊量太小或樣本太單一，測試準確率可能不穩定，無法正確反映模型的泛化能力。

4. (20 pts) Discuss methodologies for selecting relevant features in a tabular dataset for machine learning models. Highlight the importance of feature selection and how it can impact model performance. You are encouraged to consult external resources to support your arguments. Please cite any sources you refer to. (Approximately 100 words, excluding reference.)

- 過濾法：

  此方法獨立於模型，透過統計指標（如皮爾遜相關係數、卡方檢定）評估特徵與目標變數之間的關聯性，選取相關性高的特徵。過濾法計算效率高，適用於高維度資料集，但可能忽略特徵間的相互作用。

  參考資料

- 決策樹：

  決策樹透過選擇最能提升節點純度的特徵來進行資料分裂，衡量某特徵對於降低資料集熵值的貢獻，熵值越低表示資料純度越高。資訊增益大的特徵會被優先選擇。並評估資料集的不純度，不純度越低，表示資料純度越高。決策樹能夠識別出對分類結果影響最大的特徵，從而提高模型的準確性。

  參考資料

5. (20 pts) While artificial neural networks (ANNs) are versatile, they may not always be the most efficient choice for handling tabular data. Identify and describe an alternative deep learning model that is better suited for tabular datasets. Explain the rationale behind its design specifically for tabular data, including its key features and advantages. Ensure you to reference any external sources you consult. (Approximately 150 words, excluding reference.)

TabNet 是一種專為表格資料設計的深度學習模型，旨在結合決策樹的可解釋性與深度學習的高效能，其採用順序注意力機制，在每個決策步驟選擇最相關的特徵進行處理。這使模型能夠聚焦於最重要的資訊，提高學習效率和可解釋性。不同於傳統模型對所有實例應用統一的特徵選擇，TabNet 能針對每個實例動態選擇特徵。這種自適應能力使模型能夠根據不同資料點選擇最合適的特徵，從而提高預測的準確性。透過注意力機制，TabNet 能提供清晰的特徵重要性評估，揭示模型決策過程中的關鍵特徵，增強模型的透明度。其能夠在多個表格資料集上表現出優異的效能，能夠有效捕捉複雜的特徵交互，超越傳統的神經網路和決策樹模型。

參考資料
參考資料