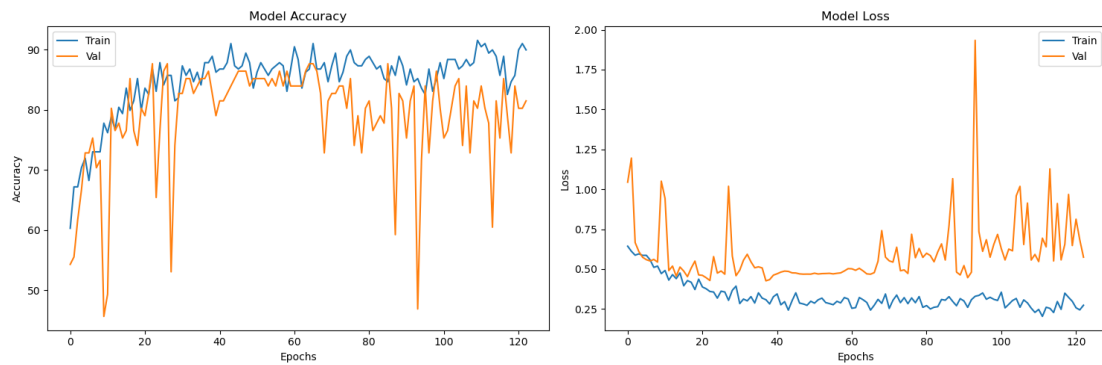


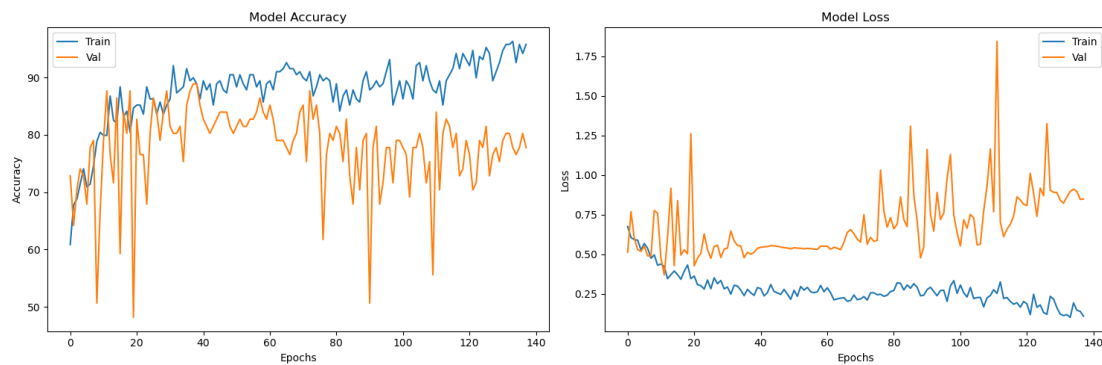
第 1 題

在本實驗中，我選擇了兩個超參數：學習率（learning rate）與隱藏層神經元數量（hidden size）。其中學習率設為 0.01、0.001、0.0001，hidden size 分別為 128、256、512，總共測試 9 組組合。每組訓練皆記錄了訓練、驗證與測試的 loss 與 accuracy，並繪製對應曲線圖與統整表格。結果顯示，當學習率為 0.001 且 hidden size 為 256 時，模型在各項指標上皆表現最佳（驗證準確率 90.12%、測試準確率 87.10%），驗證 loss 最低、收斂穩定，具良好泛化能力。

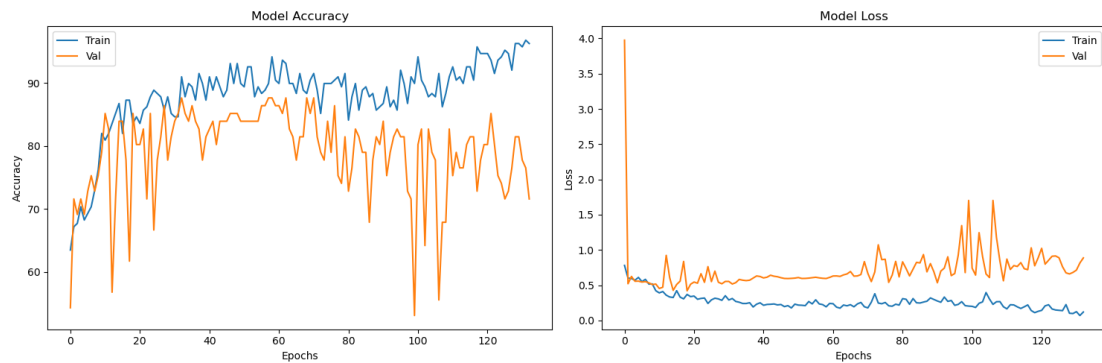
LR=0.01, Hidden=128



LR=0.01, Hidden=256

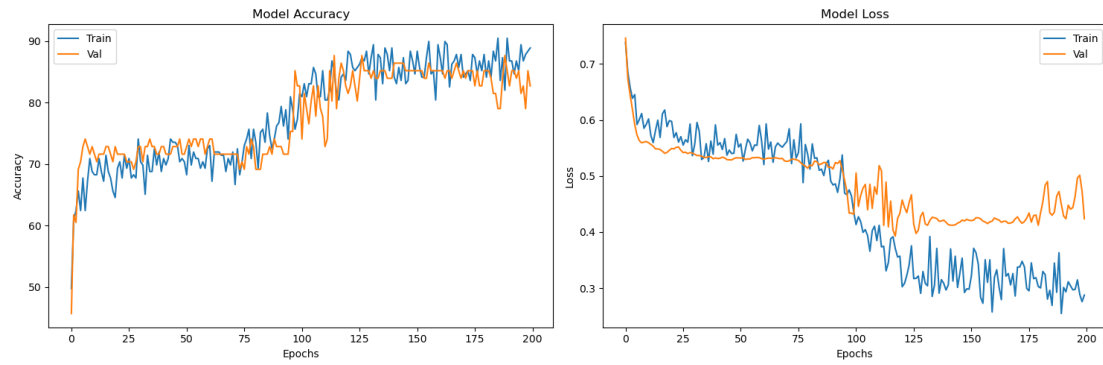


LR=0.01, Hidden=512

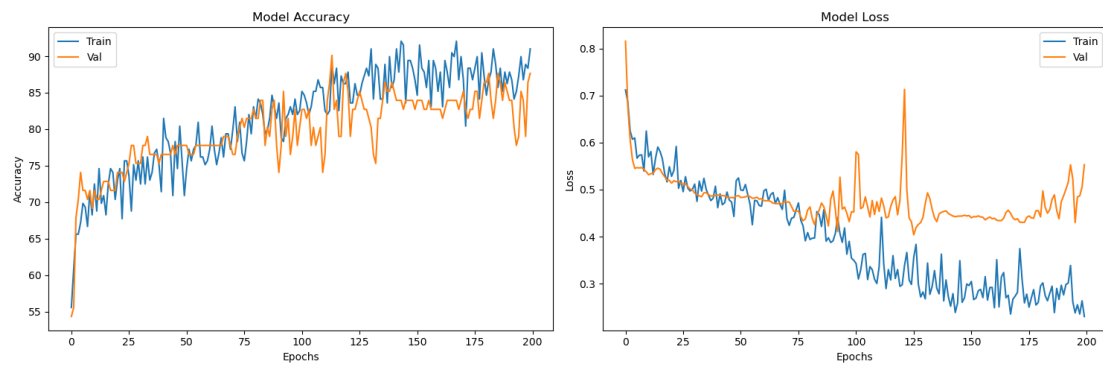


Learning Rate	Hidden Size	Best Train Loss	Best Train Acc	Best Val Loss	Best Val Acc	Best Test Loss	Best Test Acc
0.01	128	0.359596265	86.77248677	0.428285092	87.65432099	0.534456081	80.64516129
0.01	256	0.256847483	89.94708995	0.500585814	88.88888889	0.411361056	80.64516129
0.01	512	0.269182459	91.00529101	0.535897334	87.65432099	0.450579279	74.19354839

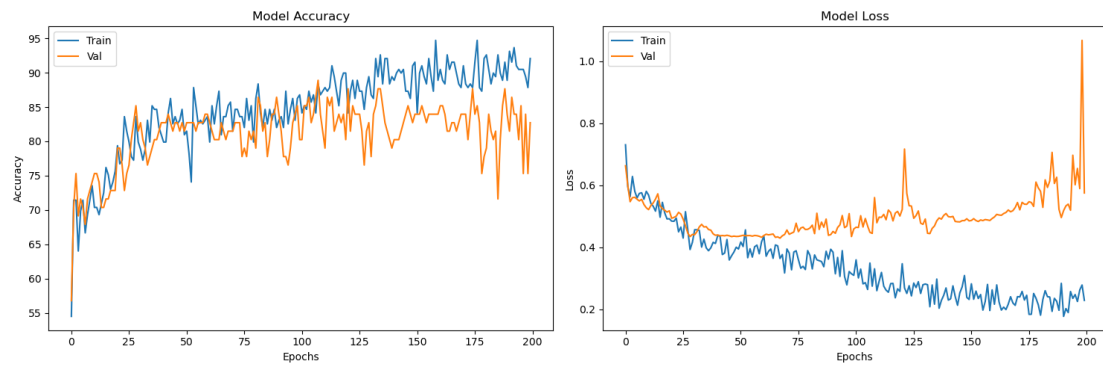
LR=0.001, Hidden=128



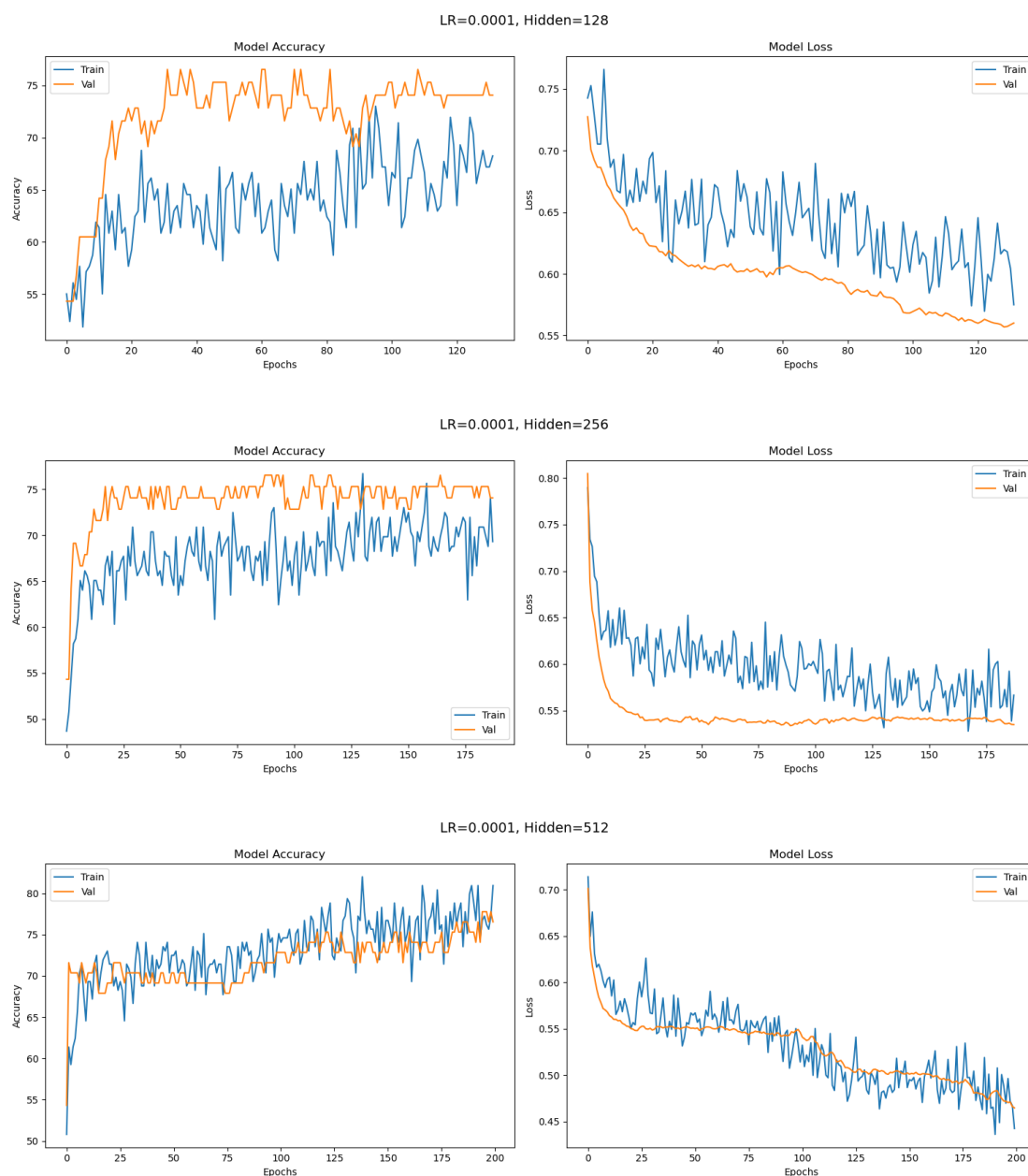
LR=0.001, Hidden=256



LR=0.001, Hidden=512



Learning Rate	Hidden Size	Best Train Loss	Best Train Acc	Best Val Loss	Best Val Acc	Best Test Loss	Best Test Acc
0.001	128	0.345607181	85.18518519	0.409348816	87.65432099	0.491940661	74.19354839
0.001	256	0.289327567	88.88888889	0.439943522	90.12345679	0.42672895	87.09677419
0.001	512	0.274458729	88.35978836	0.444610953	88.88888889	0.459039703	83.87096774



Learning Rate	Hidden Size	Best Train Loss	Best Train Acc	Best Val Loss	Best Val Acc	Best Test Loss	Best Test Acc
0.0001	128	0.637226313	65.60846561	0.60616823	76.54320988	0.637226045	67.74193548
0.0001	256	0.59916386	69.31216931	0.537325482	76.54320988	0.616156177	67.74193548
0.0001	512	0.488445769	76.71957672	0.471968383	77.77777778	0.526359837	70.96774194

第 2 題

由圖表與表格可知，學習率與 hidden size 對模型性能影響顯著。學習率 0.01 收斂速度快，但容易過擬合，驗證 loss 明顯波動；而 0.0001 則學習太慢，導致 underfitting。相較之下，0.001 為最適中設定，在多數 hidden size 組合中能穩定訓練並有效泛化。hidden size 過小模型能力不足、過大則易過擬合，以 256 最為平衡。調整這兩個超參數成功提升模型效能，最佳組合在訓練與驗證皆具高準確率且 loss 曲線平滑，顯示訓練穩定，且泛化良好。

第 3 題

準確率的差異主要是來自於模型的過擬合與資料分布不一致。在部分組合中，模型在訓練集上達到 90% 以上的準確率，但在測試集卻下降至 74%，顯示模型可能過度學習了訓練資料的特定模式或雜訊，缺乏泛化能力。此外，若訓練與測試資料在特徵分布上略有差異，也可能導致測試準確率下降。本次實驗中，我加入了 Dropout、Early Stopping 以及 Batch Normalization，以抑制過擬合並改善模型的泛化表現。

第 4 題

特徵選擇是表格型資料中提升模型效能的重要步驟，可透過多種方法進行，例如：計算特徵與目標的相關係數（如皮爾森相關）、使用隨機森林或 XGBoost 評估特徵重要性，或進行遞迴式特徵消除（RFE）。這些方法有助於刪除冗餘、無效或噪聲特徵，提升模型準確率與訓練效率。本實驗雖使用全部特徵，但搭配標準化與 BatchNorm 可減少特徵尺度不一的問題，間接達到類似效果。良好的特徵選擇能有效減少過擬合與訓練時間，是提升模型品質的關鍵。

第 5 題

對於表格型資料，TabNet 是近年發展出來更適合的深度學習模型。相較於傳統 ANN，TabNet 採用 順序式注意力機制（sequential attention），模型會自動選擇在每一層應使用的特徵子集，類似於決策樹的分支行為。這樣設計能保留特徵間的結構資訊，並強化模型解釋能力。同時，TabNet 還具備稀疏性正則化與特徵遮罩策略，可有效避免過擬合。其架構針對表格資料的特性進行最佳化，不需要額外的特徵工程，即可達到與 GBDT 相當甚至更好的效果。根據 Arik 與 Pfister（2021）的研究，TabNet 在多項表格資料任務上表現優異，成為處理 tabular data 的熱門選擇。