

	age	sex	cp	trestbps	chol	fb	restecg	thalach	exang	oldpeak	slope	ca	thal	target
age	1.000000	-0.062222	-0.103697	0.261782	0.210520	0.109847	-0.124588	-0.412624	0.111263	0.200243	-0.165360	0.254462	0.077368	-0.244798
sex	-0.062222	1.000000	-0.040197	-0.055463	-0.166885	0.042384	-0.069599	-0.058626	0.124054	0.089726	-0.038771	0.140795	0.198493	-0.283776
cp	-0.103697	-0.040197	1.000000	0.035563	-0.063592	0.065869	0.008389	0.300307	-0.428233	-0.183616	0.135174	-0.180598	-0.139765	0.425574
trestbps	0.261782	-0.055463	0.035563	1.000000	0.128444	0.170606	-0.145195	-0.056631	0.067116	0.184896	-0.126553	0.093545	0.068690	-0.173239
chol	0.210520	-0.166885	-0.063592	0.128444	1.000000	0.003430	-0.162687	-0.023753	0.063902	0.084355	-0.031929	0.068647	0.121280	-0.096773
fb	0.109847	0.042384	0.065869	0.170606	0.003430	1.000000	-0.086165	-0.014297	0.029190	0.007943	-0.056866	0.164266	-0.004972	-0.068845
restecg	-0.124588	-0.069599	0.008389	-0.145195	-0.162687	-0.086165	1.000000	0.025457	-0.089225	-0.047837	0.074982	-0.053946	-0.003377	0.101817
thalach	-0.412624	-0.058626	0.300307	-0.056631	-0.023753	-0.014297	0.025457	1.000000	-0.404349	-0.340564	0.370073	-0.205060	-0.078637	0.432687
exang	0.111263	0.124054	-0.428233	0.067116	0.063902	0.029190	-0.089225	-0.404349	1.000000	0.294308	-0.280124	0.106250	0.189253	-0.457502
oldpeak	0.200243	0.089726	-0.183616	0.184896	0.084355	0.007943	-0.047837	-0.340564	0.294308	1.000000	-0.585472	0.223375	0.200315	-0.443504
slope	-0.165360	-0.038771	0.135174	-0.126553	-0.031929	-0.056866	0.074982	0.370073	-0.280124	-0.585472	1.000000	-0.083491	-0.090606	0.363983
ca	0.254462	0.140795	-0.180598	0.093545	0.068647	0.164266	-0.053946	-0.205060	0.106250	0.223375	-0.083491	1.000000	0.136160	-0.391031
thal	0.077368	0.198493	-0.139765	0.068690	0.121280	-0.004972	-0.003377	-0.078637	0.189253	0.200315	-0.090606	0.136160	1.000000	-0.311701
target	-0.244798	-0.283776	0.425574	-0.173239	-0.096773	-0.068845	0.101817	0.432687	-0.457502	-0.443504	0.363983	-0.391031	-0.311701	1.000000

圖 1

分析特徵與 target 的相關性，決定把相關性較低的四個 features: chol, fbs, restecg, trestbps 拿掉，並得到測試準確度從 67% 上升到 87% 的結果，如圖 1。

圖 2 為準確度與損失的曲線圖，第一小題將調整超參數並與圖 2 進行比較。

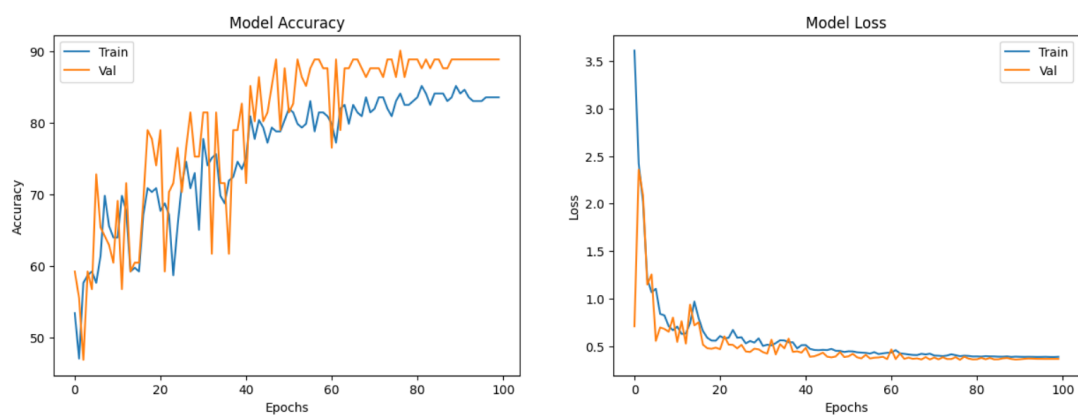


圖 2

(1)(2)

**改動 1:** 將 epoch 數從 100 調整為 200，最終的測試準確度從 87% 下降到 83%，而在驗證損失與準確度並沒有明顯的改變，因此推測此模型在 100 個 epoch 就會 overfitting，如圖 3。原先預期的結果是在 epoch 數超過 100 之後，模型的準確度會緩慢又平穩的上升，但從圖 3 看起來還是出現了蠻大的波動，驗證損失則沒有太大的變化。

再將 epoch 數從 100 調整為 50，最終的測試準確度從 87% 下降到 70%，訓練過程的損失與準確度曲線如圖 4 所示，結果仍有波動，看起來並未完全收斂，因此推斷在該模型架構下，將 epoch 數設為 80~100 數最佳的選擇。

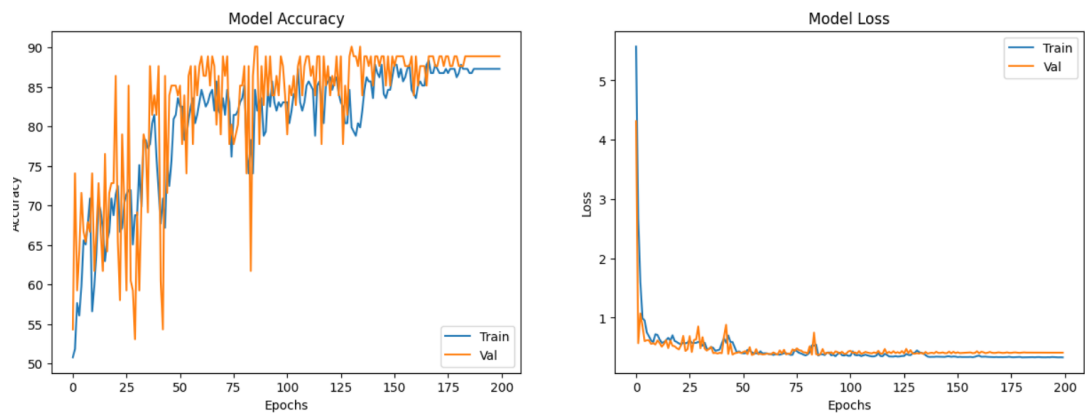


圖 3

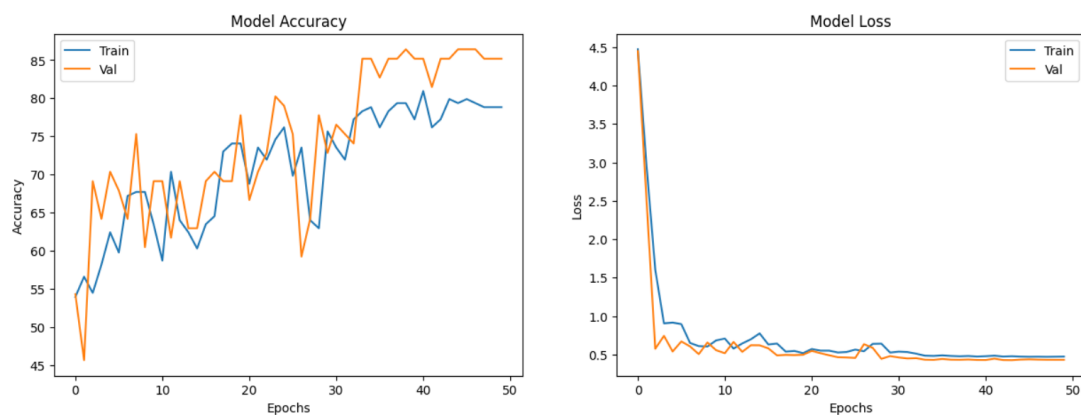


圖 4

**改動 2:** 調整 batch size，將 batch size 從 32 調整為 64，最終測試準確度卻從 87% 下降至 70%，推測若將 batch size 條大，則會使得整個訓練過程變得更佳緩慢，因此應該搭配將 epoch 數增加，以維持模型的表現。觀察圖 5 可發現準確度仍不斷波動，尚未收斂。

維持 batch size 為 64，將 epoch 數從 100 調整為 200，準確度 70% 上升到 83%，可見 batch size 增加確實需要更久的訓練時間，如圖 6，模型的訓練曲線確實有收斂。

維持 epoch 數 100，將 batch size 從 32 調整為 16，測試準確度維持在 87%，並沒有太大的變化，如圖 7，訓練結果有收斂。

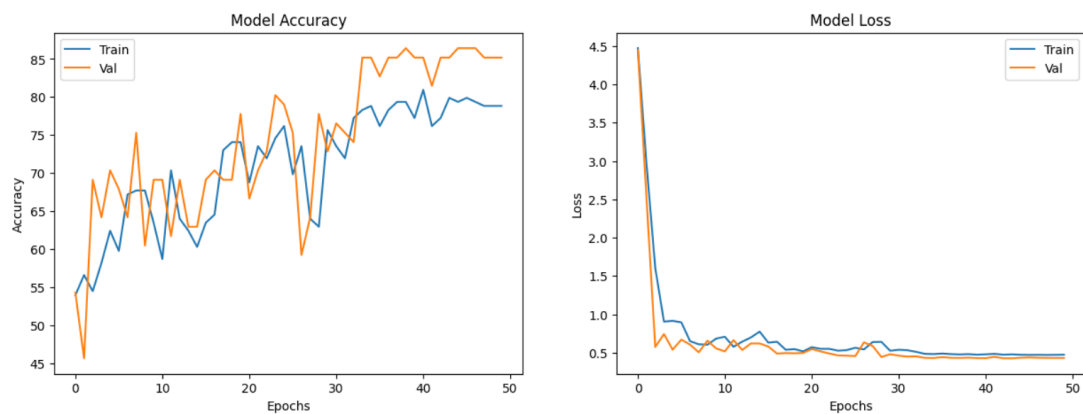


圖 5

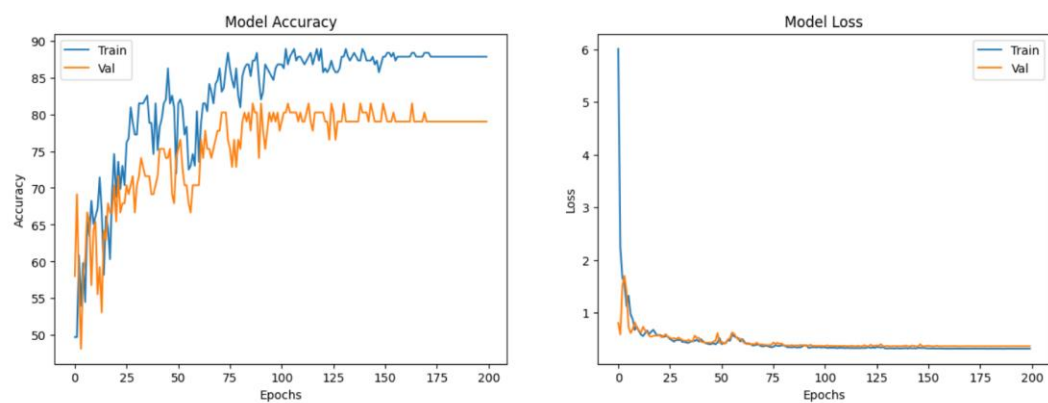


圖 6

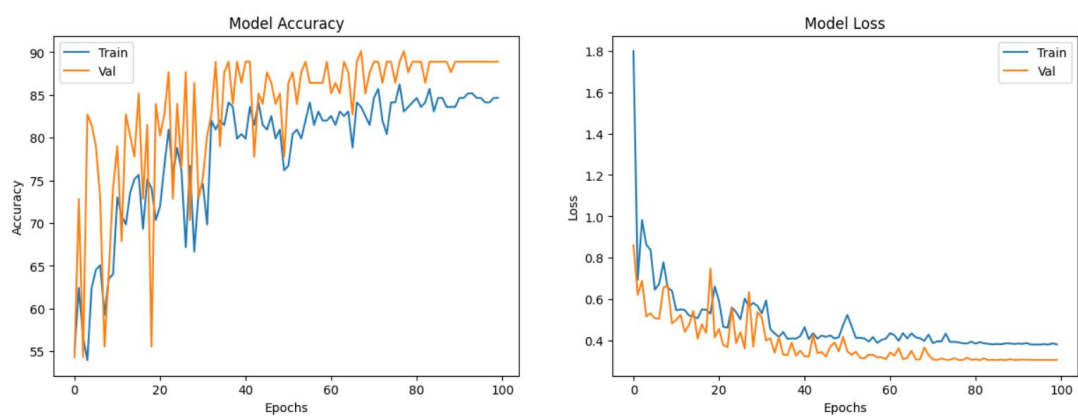


圖 7

表 1 為第一小題的總整理表

表 1

改動	訓練準確度	訓練損失	驗證準確度	驗證損失	測試準確度
Epoch = 100 Bath size = 32	85.7143%	0.3332	86.42%	0.4557	87%
Epoch = 200 Bath size = 32	88.3598%	0.3158	80.2469%	0.4188	83%
Epoch = 50 Bath size = 32	83.0688%	0.4072	85.1852%	0.4595	70%
Epoch = 100 Bath size = 64	82.5397%	0.4128	82.7160%	0.3822	70%
Epoch = 200 Bath size = 64	84.6561%	0.3824	88.8889%	0.3063	87.1%
Epoch = 100 Bath size = 16	86.7725%	0.3370	86.4198%	0.3679	87%

(3)

我發現模型在訓練集上的準確率通常比測試集高，我認為主要有三個原因。第一個是 overfitting。也就是說，模型在訓練數據上學得太好了，甚至記住了一些細節或雜訊，導致它在測試數據上表現不佳。第二個可能的原因是訓練集和測試集的數據分佈不同。如果這兩者的特徵或樣本比例差異太大，模型就難以適應測試數據，導致準確率下降。第三個原因是數據量不夠，如果訓練數據太少，模型可能學不到有代表性的特徵，而只是在記住有限的樣本。此外，也有可能有人為的因素，像是特徵選取不當或是資料不乾淨等問題。

(4)

特徵選擇的好壞直接影響模型效能。如果選擇了太多無關或冗餘的特徵，可能會導致模型學習噪音，影響準確性；相反，若忽略關鍵特徵，則可能導致模型無法準確預測。因此，選擇適當的特徵有助於提高模型泛化能力，讓它能夠更準確地適應新數據。

特徵選擇策略:

1. 過濾法：根據統計指標（如皮爾遜相關係數、卡方檢定、互資訊）來評估特徵與目標變數的關聯性，篩選出有意義的特徵。
2. 包裝法：透過特定機器學習模型（如遞歸特徵消除 RFE）來評估不同特徵組合的效能，選擇最佳組合。
3. 嵌入法：在訓練模型的過程中，同時進行特徵選擇，如 LASSO 回歸、決策樹的特徵重要性等方法。
4. 降維技術：如主成分分析（PCA），將高維度數據轉換為較少的主成分，降低維度但保留資訊量。

參考資料: Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182.

(5)

TabNet 是由 Google 提出的模型，基於注意力機制，能夠自動學習並選擇最重要的特徵來進行預測，而非像傳統 ANN 直接對所有特徵進行轉換。因此，TabNet 可以在保持高準確度的同時，仍能解釋模型的決策過程。

前面提到的 TabNet 的注意力機制並非一次性處理所有特徵，而是透過注意力機制選擇最具影響力的特徵，這樣能夠減少不必要的計算，提高學習效率。此外，傳統 ANN 需要大量超參數調整（如學習率、層數、神經元數量），而 TabNet 在較少的調整下就能取得不錯的效果。

根據文章中所提到的實驗結果，在許多表格數據應用場景（如金融風險評估、醫療診斷、行銷分析）中，梯度提升決策樹（如 XGBoost、LightGBM）通常優於傳統 ANN。而 TabNet 的設計結合了類似 GBDT 的結構，能夠學習到更好的特徵表示，提升效能。所以在表現上更勝於決策樹模型。

參考資料: Arik, S. O., & Pfister, T. (2021). TabNet: Attentive Interpretable Tabular Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(8), 6679-

6687.