



多模態資料_以阿茲海默症為例

指導老師:林真如 教授 學生:李亭萱、鄭莉蓁、蔡佳好

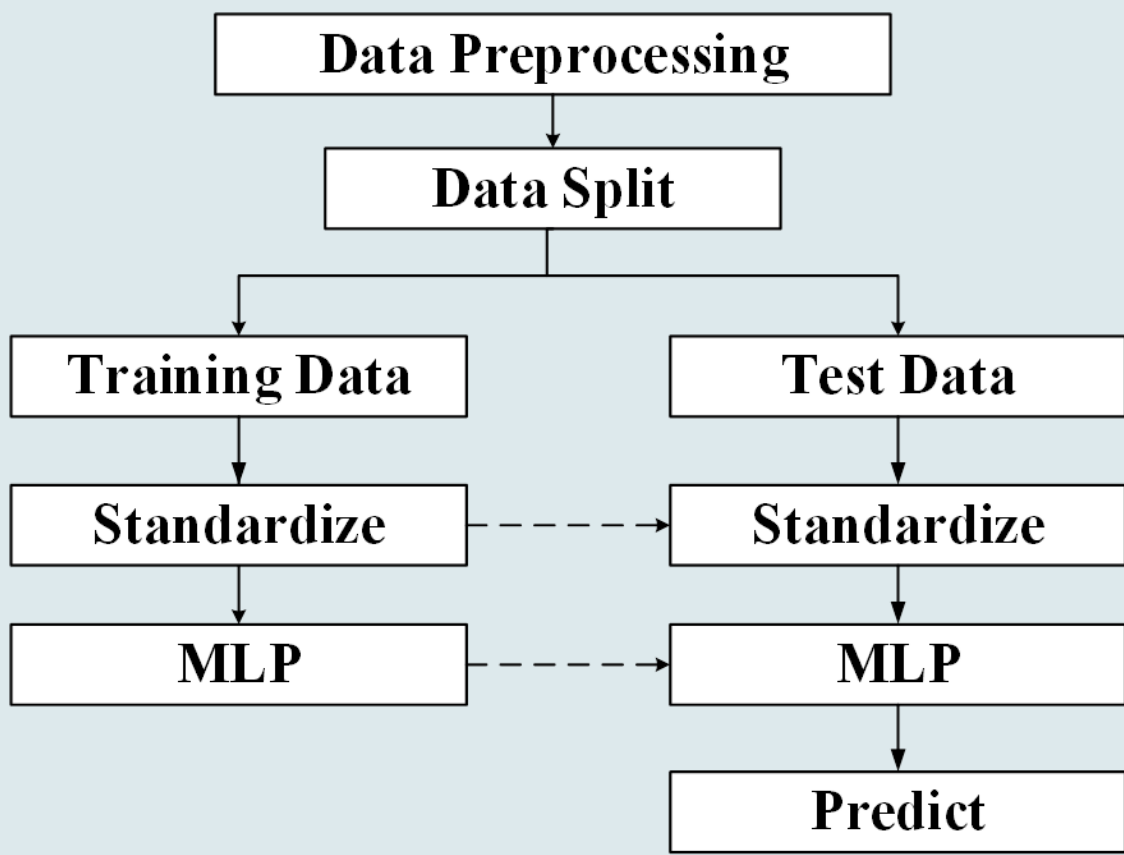
動機與目的

衛生福利部於2017年推估，台灣將在2025年邁入超高齡化社會，且於2024年統計，65歲以上失智症人口數約35萬人。對於台灣而言，失智症已是政府、社會與醫療體系需面臨的挑戰，而其中以阿茲海默症(Alzheimer's disease, AD)為主要問題。

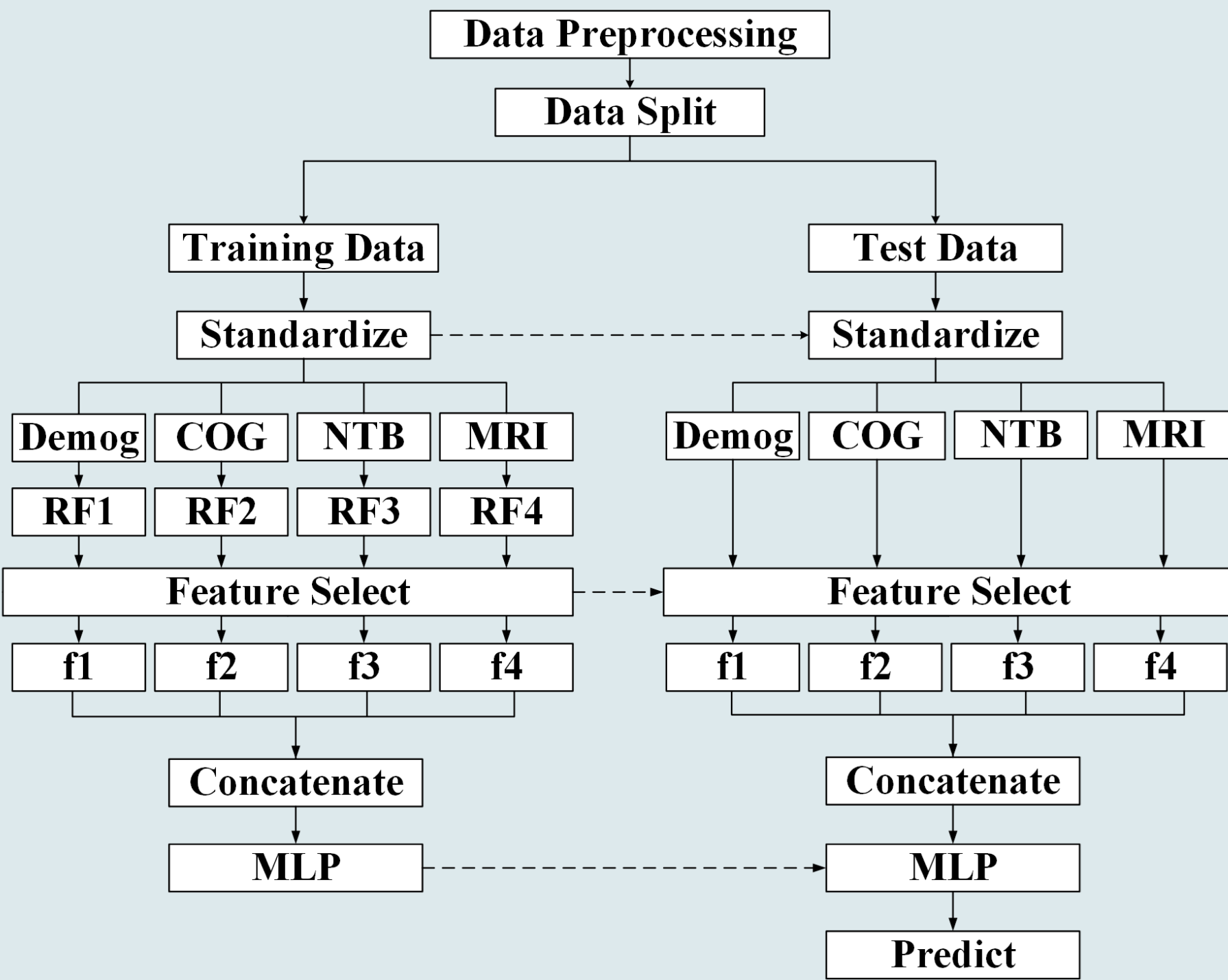
阿茲海默症需經過許多檢測才可獲得較準確的判斷，如：日常認知功能測驗、簡易心智量表、及磁振造影(Magnetic Resonance Imaging, MRI)等，這些資料型態不同，為多模態資料，因此本研究欲利用資料的獨特性與互補性，學習多模態資料的表示，融合之以達到判別病人為認知正常(Cognitively normal, CN)、輕度認知障礙(Mildly Cognitively Impaired, MCI)、與AD，提供快速且準確的工具輔助醫生診斷，及早發現疾病並治療，並藉由研究成果找出有效評估AD的重要特徵。

研究方法

本研究採用阿茲海默症神經影像倡議(Alzheimer’s Disease Neuroimaging Initiative, ADNI)資料庫中QT-PAD資料集進行分析。本研究使用人口統計數據(Demog)、認知功能測試(COG)、神經心理測驗(NTB)、腦脊隨液與MRI以分類AD。



圖一：未進行特徵篩選方法架構



圖二：內含特徵篩選方法架構

相同處

- 資料預處理：對類別型資料使用獨熱編碼(one-hot encoding)
- 資料分割：訓練：驗證：測試=64：16：20
- 標準化：min-max
- K折交叉驗證

相異處

- 保留重要性不為0
- 保留重要性累積至90%
- 保留重要性累積至80%
- 保留重要性累積至60%

研究數據

表一：各類狀態樣本數

類別	CN	MCI	AD	總數
樣本數	203	230	84	517

表二：各模態特徵說明

模態	特徵說明	特徵數
D e m o g	年齡、性別、教育程度、婚姻狀態、BMI、APOE4、基礎症狀檢核、病史。 腦脊隨液：乙型類澱粉蛋白(ABETA)、tau蛋白(TAU)和磷酸化tau蛋白(PTAU)	59
C O G	• 阿茲海默症評估量表 • 蒙特婁認知評估 • 簡易心智量表 • 老年憂鬱症量表 • 功能評估問卷 • 臨床失智評估量表	7
N T B	• 日常認知功能測驗 • Rey聽覺語言學習測驗 • 波士頓命名測驗 • 畫鐘測驗	49
M R I	• 表面積 • 皮質下體積 • 皮質體積 • 皮質厚度平均值 • 皮質厚度標準差	313

結論與未來研究方向

在多模態分類中，變數過多引發高維度問題會降低模型準確率與效率，因此特徵篩選和模態融合策略十分重要。本研究利用隨機森林和MLP篩選重要特徵，探討不同篩選策略對模型表現的影響，並因應預測類別不均衡，採用宏觀平均值作為評估標準。

未來研究方向可以使用Shapley Additive explanation (SHAP)反推計算各個特徵對模型預測的貢獻性，並與隨機森林所得的特徵重要性排序進行比較，透過兩種方法篩選出的關鍵特徵，將為後續AD診斷提供重要依據。

參數設定

表三：隨機森林參數設定

隨機森林 模型參數設定	特徵篩選方式				
	原始資料	保留非0	累積至90%	累積至80%	累積至60%
亂數種子	42				
決策樹	100				
最大深度	3		5		
最小節點	5				
最小分裂樣本數	5		-		
CPU	-1				

表四：MLP參數設定

MLP 模型訓練參數	特徵篩選方式				
	原始資料	保留非0	累積至90%	累積至80%	累積至60%
亂數種子	42				
隱藏層、輸出層	(255,170,85,30,3)	(225,170,85,30,3)	(100,50,25,3)	(130,60,35,3)	(100,50,25,3)
丟棄率 (Dropout rate)	(0.3,0.5,0.7,0.75)	(0.3,0.5,0.7,0.75)	(0.7,0.7,0.7)	(0.5,0.7,0.8)	(0.7,0.7)
激活函數	隱藏層：ReLU 輸出層：Softmax				
優化器	自適應矩估計(Adaptive Moment Estimation, Adam)				
損失函數	分類交叉熵(Categorical Cross Entropy)				
模型評估指標	分類準確率(Categorical accuracy)				
迭代次數	145	145	200	200	200

研究結果

表五：分類結果

		宏觀平均值	
		平均值	標準差
原始資料	精確率	0.768	0.072
	召回率	0.806	0.049
	F ₁ -score	0.780	0.063
	特異度	0.878	0.033
保留非0	精確率	0.805	0.037
	召回率	0.802	0.041
	F ₁ -score	0.800	0.038
	特異度	0.894	0.019
累積至90%	精確率	0.771	0.034
	召回率	0.778	0.030
	F ₁ -score	0.767	0.036
	特異度	0.876	0.012
累積至80%	精確率	0.822	0.039
	召回率	0.822	0.019
	F ₁ -score	0.816	0.022
	特異度	0.899	0.010
累積至60%	精確率	0.790	0.064
	召回率	0.794	0.042
	F ₁ -score	0.788	0.055
	特異度	0.886	0.026

- CDRSB、ABETA和PTAU三個特徵在四種策略中均為關鍵特徵，與認知功能測試和罹患阿茲海默症風險相關。
- 前五名中的ADAS13、MMSE（認知功能測試）和EcogSPTotal（神經心理測驗），反映出此兩種模態為重要的判斷項目。

- 特徵重要性累積至80%時，模型表現最佳，精確率與召回率皆為0.822，F₁-score為0.816，特異度為0.899。
- 標準差愈低，表示模型預測效果佳且穩定性高。
- 適度篩選特徵能提升模型的效率和預測準確性。

表六：各特徵篩選策略之特徵重要性

	特徵名稱	特徵重要性
保留非0	CDRSB	0.4582
	ABETA	0.1919
	PTAU	0.1494
	ADAS13	0.1406
	EcogSPTotal	0.1393
累積至80%	CDRSB	0.4993
	ABETA	0.1759
	PTAU	0.1483
	ADAS13	0.1287
	MMSE	0.1253
累積至60%	CDRSB	0.4993
	ABETA	0.1759
	PTAU	0.1483
	ADNI_MEM	0.1251
	EcogSPTotal	0.1092