



基於多模態分類之腦中風預測分析

指導教授:林真如 博士 學生:劉怡杉、翁郁雯、黃新韻

研究動機與目的

中風是一種突發性的腦血管疾病，嚴重將導致死亡。目前診斷方式仍依賴高精度影像，檢測成本高，缺乏預測性診斷。因此，本研究結合機器學習與多模態資料，建立**多種預測模型**，並探討**不同資料組合與融合策略**效能，以達到**日常監測**與提前預防腦中風疾病的目的。

研究方法

資料收集及資料預處理

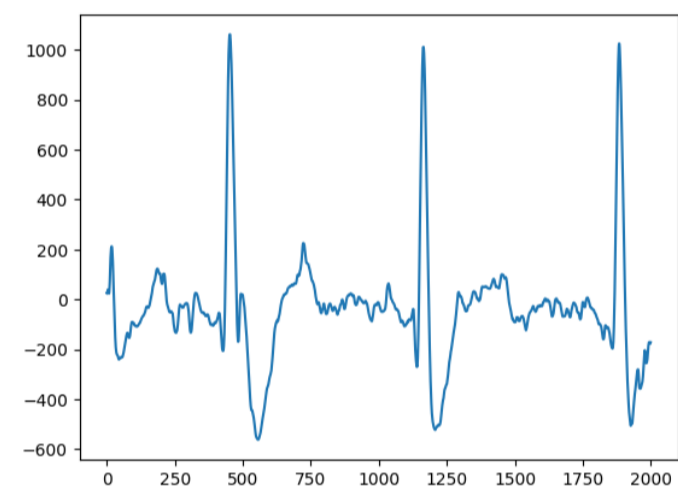
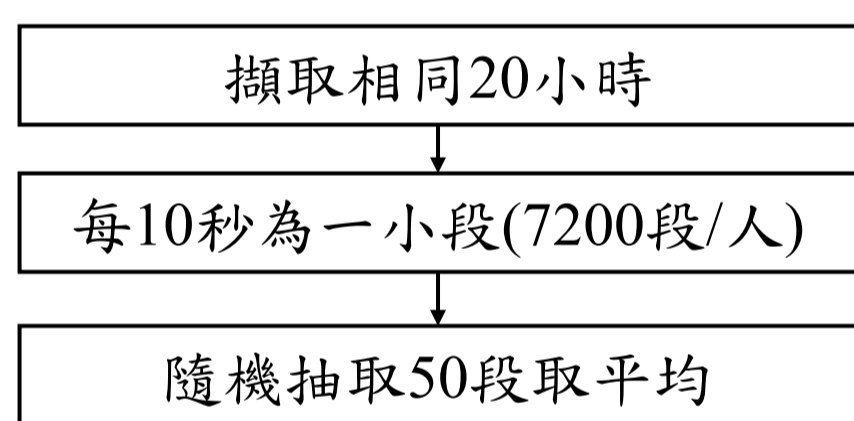
➢ PhysioNet 公開資料(非中風：51人；中風：45人)

資料型態	結構化			非結構化
資料種類	基本資料	步行紀錄	血壓數值	心電圖
實驗方式	x	12分鐘實驗	24小時量測/ 8種實驗	24小時
特徵數量	37個	11個	92個	66個

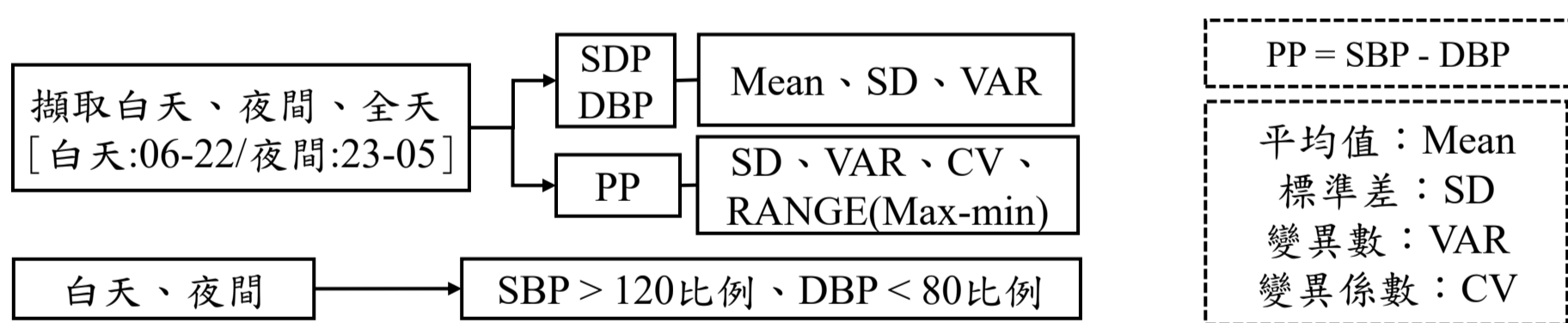
1. 欄位缺失值 => 平均值補值
2. 資料不平衡 => 類別權重(Class Weights)
3. 資料欄位數值差異 => Min-max正規化

特徵工程

- 心電圖**
 - 使用python Neurokit2工具包
 - 進行心率變異性分析 => 時域、頻域、非線性域
 - 萃取 66 項特徵值

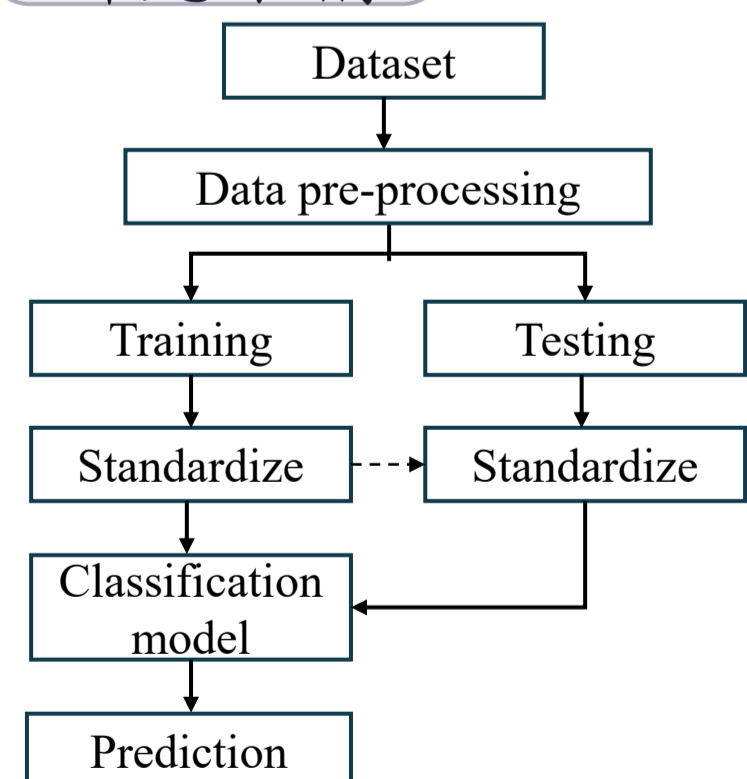


- 血壓數值**
 - 收縮壓(SBP)、舒張壓(DBP)、脈壓(PP)
 - 加入相關統計特徵，共新增 26 個特徵



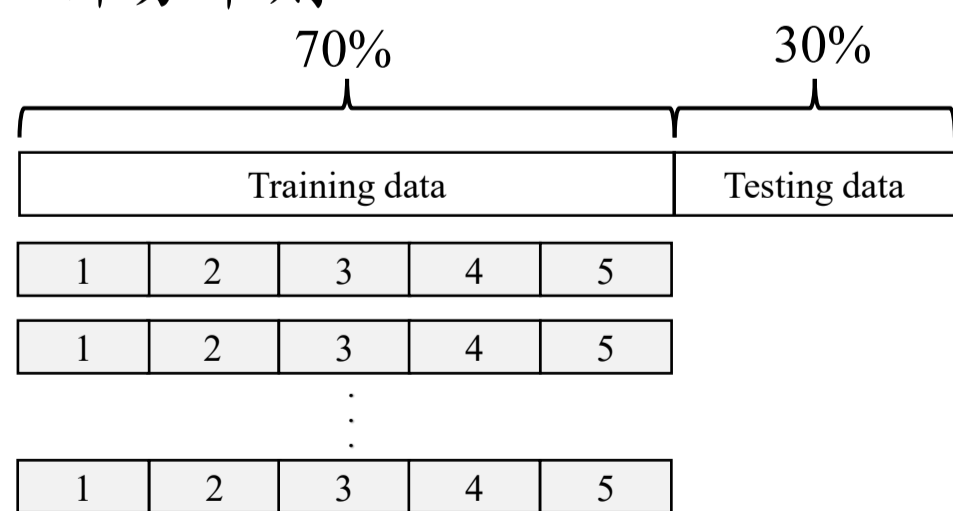
模型與融合策略分析

研究架構

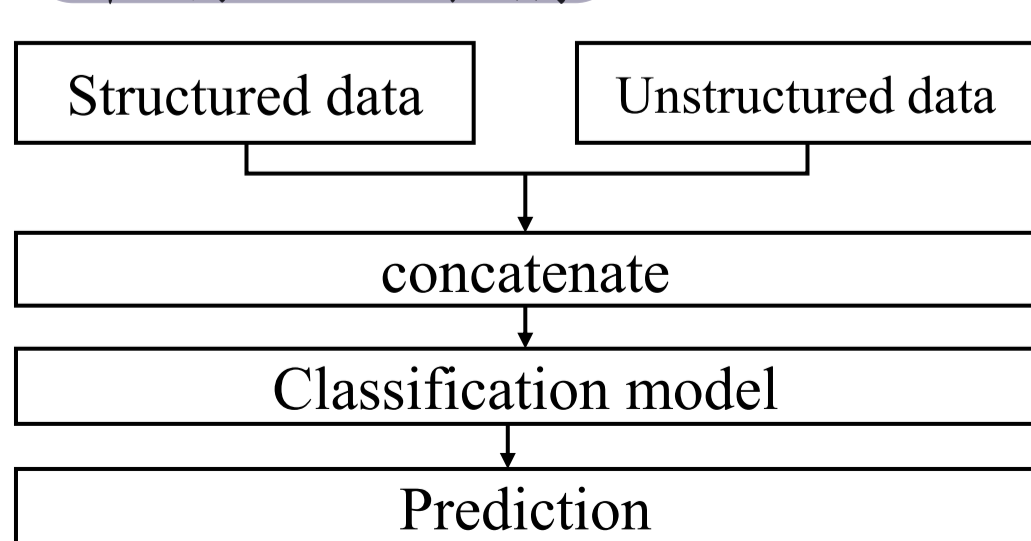


交叉驗證

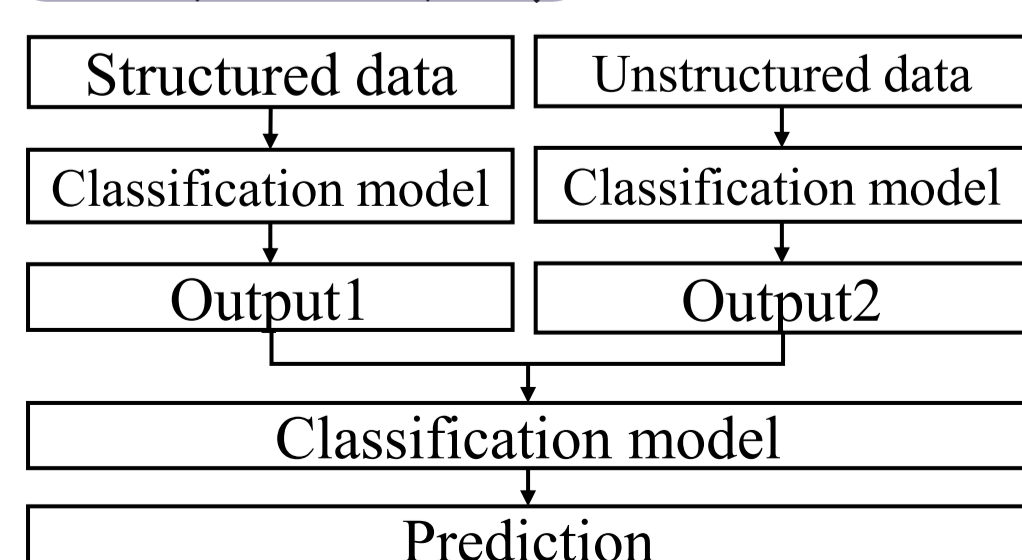
- 網格搜索-最佳化參數
- 分層五折交叉驗證
- 評分準則：F1-score



早期融合架構



晚期融合架構



▲ 投票法(voting)：

- sum_votes = Output1(步行) + Output2(血壓) + Output3(心電圖)
- 若sum_votes >= 2 -> 判為腦中風

研究結果

	屬性	資料組合	Model	Accuracy	Precision	Recall	F1	AUC	PR	
單模態資料	結構化	步行紀錄	RF	0.72	0.73	0.72	0.72	0.85	0.90	
			SVM	0.70	0.73	0.70	0.68	0.59	0.71	
			XGBoost	0.62	0.63	0.62	0.62	0.79	0.85	
		血壓數值	RF	0.86	0.87	0.86	0.86	0.88	0.85	
			SVM	0.58	0.61	0.58	0.52	0.62	0.65	
			XGBoost	0.79	0.80	0.79	0.79	0.82	0.82	
	步行+血壓	RF	0.76	0.77	0.76	0.76	0.85	0.88		
		SVM	0.50	0.50	0.50	0.50	0.60	0.56		
		XGBoost	0.82	0.81	0.82	0.82	0.90	0.88		
	非結構化	心電圖	RF	0.76	0.76	0.76	0.76	0.89	0.89	
			SVM	0.50	0.50	0.50	0.49	0.41	0.45	
			XGBoost	0.76	0.77	0.76	0.76	0.85	0.88	
多模態資料	結構 + 非結構化	步行紀錄 + 血壓數值 + 心電圖	早期融合	RF	0.79	0.85	0.79	0.78	0.88	0.90
			SVM	0.65	0.65	0.65	0.65	0.60	0.55	
			XGBoost	0.83	0.84	0.83	0.83	0.91	0.91	
	晚期融合	RF	0.9	0.92	0.9	0.9	0.96	0.91		

➢ 分類器比較

- 單模態：RF 表現佳，Accuracy達0.86；
- 多模態：XGBoost 表現佳，Accuracy達0.83
- SVM表現最差，無法有效進行二元分類

➢ 單模態資料建模

- 血壓數值：Accuracy達0.86，預測模型效果最佳

➢ 多模態資料建模-早期融合 & 晚期融合

- 早期融合Accuracy達0.83；晚期融合Accuracy達0.9
- 晚期融合(多數決投票)能提升分類結果

➢ 單模態 & 多模態資料建模

- 多模態融合後，相比最佳單模態的準確率和穩定度提升
- 驗證多模態資料於醫療領域仍有一定研究價值

結論與未來展望

5種資料組合 + 2種融合策略 => 腦中風預測

➢ 研究發現

- 預測腦中風疾病：多模態資料 + 晚期融合
- 使用樹狀結構模型對於疾病預測有較佳效果
- 不論是用於日常監測或是醫療監測，若受限於醫療設備，則可使用**血壓數值**先進行判斷腦中風情況

➢ 未來展望

- 優化資料預處理方法、模型參數，提升分類的預測準確性
- 嘗試更多不同分類器建立模型
- 結合更多生理訊號與臨床資料，強化模型臨床應用價值