

基於多模態聚類與特徵篩選之中醫虛證辨識研究

A Study on Deficiency Syndrome Identification in Traditional Chinese Medicine Based on Multimodal Clustering and Feature Selection

研究生：資訊工程系 林宗瑋

指導教授：資訊工程系 張傳育 特聘教授

摘要

人體生理機能衰退常於初期不易察覺，許多人在尚未形成明確疾病診斷前，可能已出現體力下降、呼吸能力減弱、精神疲倦或活動耐受度降低等健康變化。在中醫理論中，此類人體機能低下、活動力不足與恢復能力下降等狀態，可對應於虛證相關表現。

既有虛證量化研究多以病人主訴、醫師觀察所得之症狀與體徵，相較之下，本研究使用客觀資料，透過儀器量測或標準化測試程序取得數值資料，以減少單純依賴症狀描述所產生之個人表達與判斷差異，建立一套結合多模態客觀量測資料、聚類分析、特徵篩選、虛證分數計算之虛證客觀化分析流程，研究資料包含肺功能、語音聲學、行動功能及心率血壓等特徵，並以 FAMvC 多模態聚類方法進行聚類，FAMvC 為一種多視圖聚類方法，其核心在於分離不同資料來源間的一致與不一致性資訊，以提升異質資料整合的穩健性與分群表現。接著，依據醫師臨床意義認定後之群體標籤，採用 Sequential Floating Forward Selection (SFFS) 進行特徵篩選。最後，針對非虛群與虛群進行 Mann - Whitney U Test，分析虛與非虛群特徵之差異，建立整體虛證分數，並以 ROC 曲線評估其區辨能力。

研究結果顯示，聚類形成之群體經醫師判讀後，可分別解釋為非虛群、過渡群與虛群，呈現由非虛至虛之程度變化。原始 20 項特徵經篩選後保留 17 項代表性特徵。整體虛證分數於非虛群與虛群、非虛群與過渡群，以及過渡群與虛群之 AUC 分別為 0.987、0.774 與 0.908。依據相鄰群體之 ROC 切點，本研究提出探索性分數區間：低於 0.698 者較接近非虛狀態，0.698 至 0.871 者較接近過渡狀態，高於 0.871 者則較接近虛狀態。研究結果顯示，多模態客觀生理與功能資料可用於探索虛證相關群體、辨識代表性特徵，並建立反映由非虛至虛程度變化之連續量化分數，具有應用於虛證客觀化評估與健康狀態追蹤之潛力。。

研究方法

本研究首先收集多模態客觀資料，包含肺功能、聲音聲學、步態影片、TUG 影片，以及心率與血壓資料，並針對不同資料型態分別進行特徵提取，完成特徵提取後，本研究先嘗試將資料分為二群，以觀察資料是否能直接呈現非虛與虛之分化。然而，初步二群聚類結果顯示，部分受試者之特徵表現介於非虛與虛之間，並未完全呈現明確的二分結構。此結果顯示非虛與虛之間可能存在過渡狀態，因此本研究進一步將聚類數調整為三群，在三群聚類結果中，本研究進一步由醫師根據各群之特徵表現與臨床意義進行審查，並將三個群組分別標記為非虛群、過渡群與虛群。此方式可將資料驅動之聚類結果與臨床判斷相結合，使分群結果更具臨床解釋性，基於醫師審查後之三群結果，本研 Sequential Forward Floating Selection (SFFS) 進行關鍵特徵子集篩選，以找出能夠區分非虛、過渡與虛三群之重要特徵，為建立整體虛證分數，本研究進一步針對非虛群與虛群進行 Mann - Whitney U test，篩選兩群之間具有顯著差異的關鍵特徵並計算出整體虛證分數，用以量化受試者之虛證程度，作為臨床輔助評估之參考依據。本研究提出的分析流程如圖 1 所示。

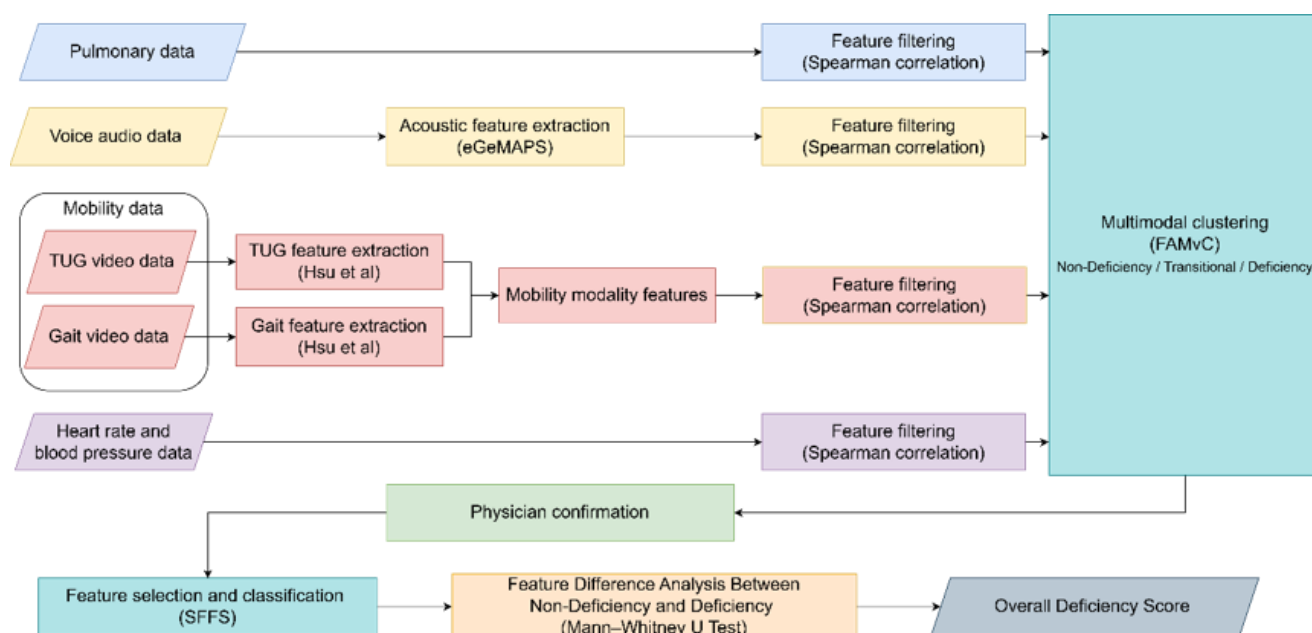


圖 1. 分析流程圖

實驗結果

本研究資料來源為台灣中部大林慈濟醫院中醫部所提供之多種生理與行為模態資料，以反映受試者於不同功能面向的整體狀態，其中，肺功能模態包含 FEV1、PEF 及 FEV1% predicted 等儀器量測數值；聲音模態為錄音設備所錄製之 5 秒、16-bit、44 kHz 音檔；行動功能模態包含錄影設備所取得之 640 × 480、30 fps TUG 與步態影片；心率血壓模態則包含 HR、SYS、DIA 等儀器量測數值，為

進一步評估聚類結果是否符合臨床判斷，實驗同時蒐集醫師對各樣本之虛與非虛判斷結果，作為聚類結果之事後對照依據，其分布如表 1 所示

表 1. 實驗資料集樣本分布

	項目	虛證樣本 (筆)	非虛證樣本 (筆)	合計 (筆)
性別	Male	23	51	74 (29.1%)
	Female	118	62	180 (70.9%)
合計 (筆)		141	113	254 (100.0%)
年齡	<65 years	59 (41.8%)	69 (61.1%)	128 (50.4%)
	65–74 years	65 (46.1%)	37 (32.7%)	102 (40.2%)
	≥75 years	17 (12.1%)	7 (6.2%)	24 (9.4%)
合計 (筆)		141 (100.0%)	113 (100.0%)	254 (100.0%)

因初步二群聚類結果顯示，部分受試者之特徵表現介於非虛與虛之間，並未完全呈現明確的二分結構，表示單純二分設定可能不足以完整描述資料中虛與非虛之間的樣本狀態故本研究後續進一步採用三聚類設定進行分析嘗試，其結果如圖 2 所示。

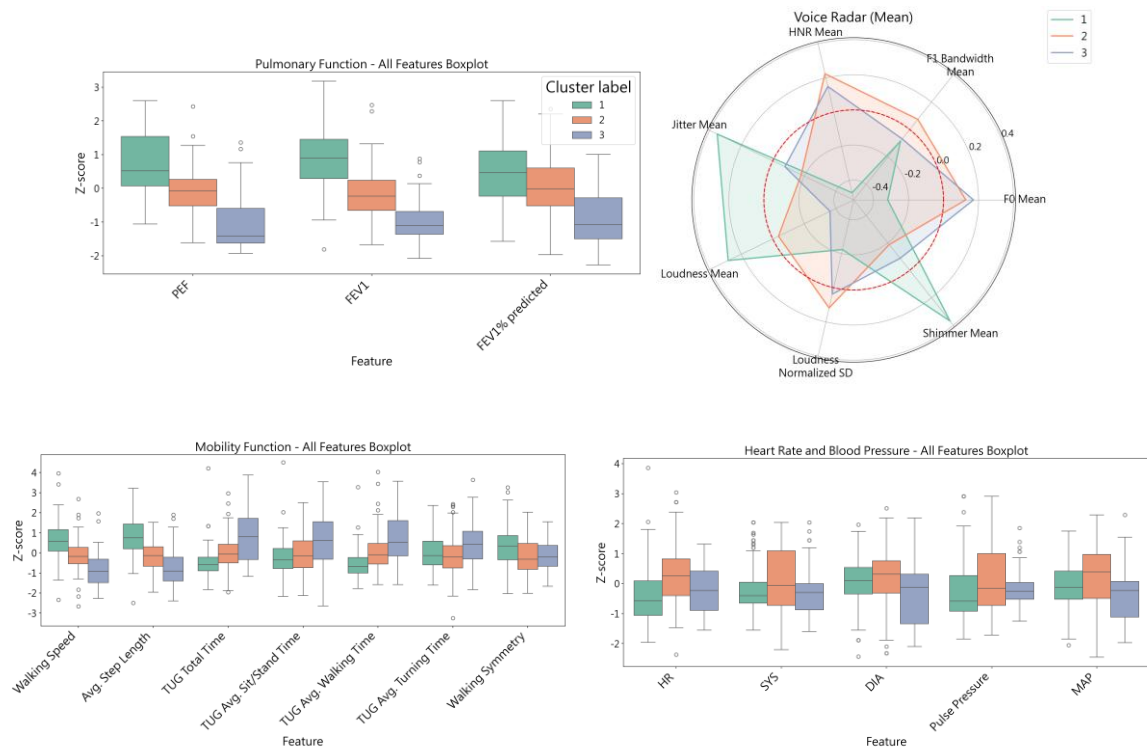


圖 2. 各模態三聚類特徵分布圖

三聚類結果將樣本分為 Cluster 1：77 筆、Cluster 2：130 筆 與 Cluster 3：47 筆，綜合各模態

數值分布結果可知，三聚類所得群體在肺功能、聲音及行動能力等模態上呈現可解釋之差異。其中，Cluster 1 與 Cluster 3 分別呈現較佳與較弱之功能表現，而 Cluster 2 在特徵上多介於兩者之間，顯示其可能反映二分設定下較難明確歸屬之中介型樣本，基於此，醫師將 Cluster 1、Cluster 2 與 Cluster 3 分別認定為非虛群、過渡群與虛群。

為降低特徵冗餘並找出具判別力且穩定的特徵子集，本研究將整體資料劃分為 80% 訓練集 (development set) 與 20% 獨立測試集 (independent test set)，其中訓練集再分為 60% 特徵篩選資料與 20% 特徵子集評估資料，本研究於 60% 特徵篩選資料中進行 10 次重複五折交叉驗證，並於各折執行 SFFS，共取得 50 組最佳特徵子集。依各特徵於 50 組子集中的出現比例計算選取頻率，並設定不同頻率門檻形成候選特徵子集。各候選子集以五折交叉驗證評估分類效能，選出最佳門檻與對應特徵子集，再於另外 20% 評估資料中確認其泛化表現，且不再依評估結果調整門檻，完成特徵子集確認後，將上述資料合併為完整 80% 訓練集，並以 Mann - Whitney U 檢定比較非虛證組與虛證組之特徵差異。依統計顯著性計算特徵權重，結合特徵方向與標準化特徵指標，建構個別受試者之整體虛證分數，最後，固定訓練集中所確定之特徵子集、參考值、特徵方向與權重，直接套用於保留的 20% 獨立測試集，並利用 ROC 曲線及曲線下面積 (area under the curve, AUC) 評估虛證分數對不同虛證程度之區辨能力，其流程如

圖 3 所示，各群別於資料切分後之樣本數如表 2 資料集之樣本分布及錯誤! 找不到參照來源。所示。

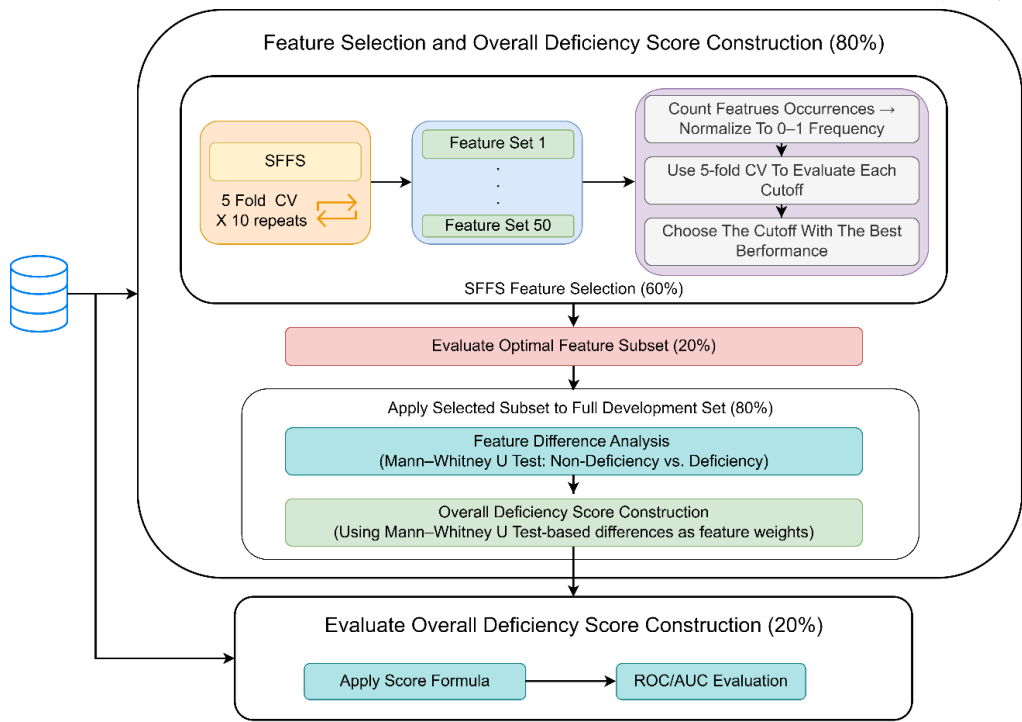


圖 3. 特徵篩選及虛證分數計算流程圖

表 2 資料集之樣本分布

類別	訓練集 (80%)	獨立測試集 (20%)	合計(100%)
非虛	57	15	72
過渡	104	26	130
虛	42	10	52
合計	203	51	254

表 3 訓練集內部資料分布

類別	特徵篩選子集	內部驗證子集	訓練集
非虛	43	14	57
過渡	78	26	104
虛	31	11	42
合計	152	51	203

於不同特徵選取頻率門檻下，可形成不同的候選特徵子集。本研究分別以五折交叉驗證（5-fold Cross Validation）評估各候選特徵子集於訓練資料集上的分類效能，並以 Macro F1 作為評估指標。結果顯示，當 frequency threshold 設定為 0.42 時，模型可獲得最高的 Macro F1，此時共保留 17 個特徵，因此選定 0.42 為最佳特徵選取頻率門檻，當門檻設定較低時，雖可保留較多特徵，但可

能同時納入穩定性或判別能力較低的特徵，使模型效能未達最佳；相反地，當門檻設定過高時，僅保留出現頻率較高的特徵，可能排除部分仍具有分類貢獻的資訊，進而造成模型效能下降。此結果顯示，適當的特徵選取頻率門檻可在特徵穩定性與分類效能之間取得較佳平衡，其不同頻率門檻下之模型分類效能及篩選出的特徵如圖 4 及表 4 所示。

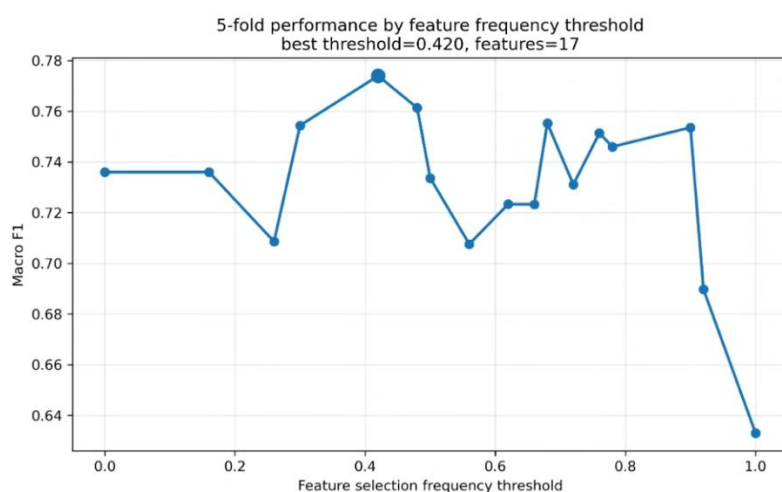


圖 4 不同特徵出現頻率門檻下之模型分類效能

表 4 Frequency threshold = 0.42 下之最終特徵子集

特徵名稱	模態	出現次數	頻率閾值
PEF	肺功能	50	1
FEV1	肺功能	46	0.92
HR	心率&血壓	45	0.9
FEV1% Predicted	肺功能	39	0.78
SYS	心率&血壓	38	0.76
TUG Total Time	行動功能	38	0.76
TUG Avg. Sit/Stand Time	行動功能	36	0.72
Pulse Pressure	心率&血壓	34	0.68
TUG Avg. Turning Time	行動功能	34	0.68
Walking Speed	行動功能	33	0.66
Loudness Normalized SD	聲音	33	0.66
Mean Shimmer	聲音	31	0.62
Avg. Step Length	行動功能	28	0.56
Mean HNR	聲音	25	0.5

Mean Jitter	聲音	25	0.5
DIA	心率&血壓	24	0.48
Walking Symmetry	行動功能	21	0.42

接著本研究採用 Mann–Whitney U Test 比較非虛群與虛群於各特徵之分布差異，並使用 Benjamini–Hochberg 方法進行 FDR 校正，以降低多重比較所造成之偽陽性風險。檢定後達顯著差異之特徵進一步依據虛群相對於非虛群之平均值差異判定方向；若虛群平均值高於非虛群，則標示為「虛群較高」，反之則標示為「虛群較低」，並將各特徵經 Mann – Whitney U 檢定所得之 p 值進行 FDR 校正，再以 $-\log_{10}(\text{FDR p-value})$ 轉換並正規化，作為各特徵之相對權重。FDR 校正後 p 值越小的特徵，其權重越高，因此肺功能與部分行動功能特徵對虛證分數的影響較大，其結果如表 5 所示。

表 5 非虛群與虛群之 Mann–Whitney U Test 結果

特徵名稱	模態	非虛/虛 平均數	FDR P- VALUE	權重	方向
PEF	肺功能	424.82 / 176.43	2.67E-12	0.209	虛群較低
FEV1	肺功能	2.29 / 1.26	2.55E-11	0.192	虛群較低
FEV1% Predicted	肺功能	0.90 / 0.63	6.39E-09	0.148	虛群較低
Avg. Step Length	行動功能	61.51 / 52.87	1.73E-06	0.104	虛群較低
Walking Speed	行動功能	130.77 / 105.50	3.45E-06	0.099	虛群較低
TUG Total Time	行動功能	10.90 / 13.71	1.11E-05	0.090	虛群較高
TUG Avg. Sit/Stand Time	行動功能	1.61 / 1.92	0.0008	0.055	虛群較高
Loudness Normalized SD	聲音	0.169 / 0.257	0.0177	0.032	虛群較高
Walking Symmetry	行動功能	0.155 / 0.109	0.0477	0.024	虛群較低
Pulse Pressure	心率&血壓	45.00 / 47.50	0.1582	0.014	虛群較高
SYS	行動功能	134.98 / 138.17	0.2693	0.010	虛群較高
TUG Avg. Turning Time	聲音	1.23 / 1.29	0.2693	0.010	虛群較高
Mean Jitter	心率&血壓	0.0141 / 0.0136	0.6609	0.003	虛群較高
HR	聲音	74.75 / 72.71	0.6609	0.003	虛群較低
Mean HNR	聲音	7.82 / 8.62	0.6944	0.003	虛群較低
Mean Shimmer	心率&血壓	0.945 / 0.781	0.6992	0.003	虛群較高
DIA	心率&血壓	89.98 / 90.67	0.923406	0.001	虛群較高

最後，透過關鍵特徵及其對應權重計算個別受試者之整體虛證分數，並比較非虛群、過渡群與虛群之分數分布，以分析整體虛證分數與不同虛證狀態之關係，進一步提出三群之建議分數區間。三群整體虛證分數之分布情形如圖5所示。

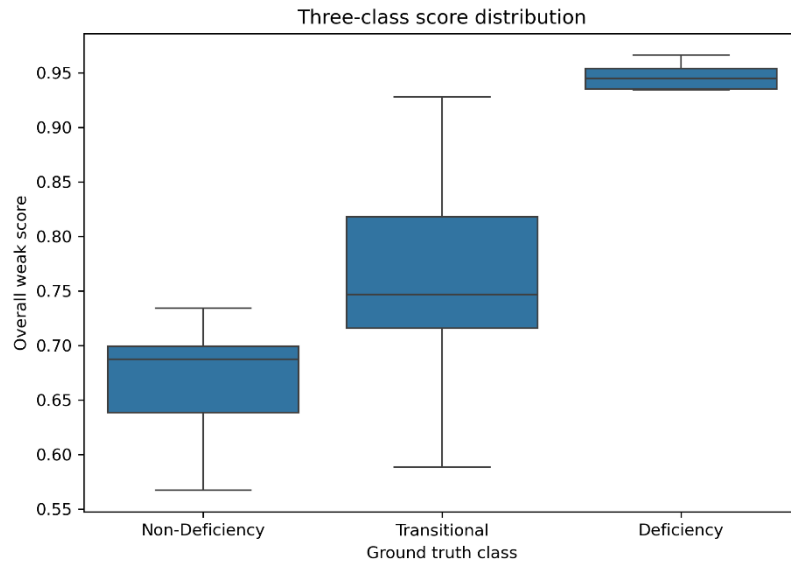


圖5 非虛群、過渡群與虛群之整體虛證分數分布

非虛群之整體虛證分數主要分布於較低區域，過渡群之分數主要位於中間區域，而虛群之分數整體較高，呈現由非虛、過渡至虛逐漸增加之分布趨勢，進一步經 Kruskal–Wallis Test 比較三群整體虛證分數，結果顯示三群間具有顯著差異 ($p = 5.36 \times 10^{-6}$)。此結果表示，整體虛證分數之分布與醫師對非虛、過渡及虛之三群判斷具有一致趨勢，顯示該分數可反映不同虛證狀態間之整體差異，為進一步觀察整體虛證分數對三群間之區辨能力，本研究針對非虛群與過渡群、非虛群與虛群，以及過渡群與虛群分別進行 pairwise ROC 分析，結果如圖-6 所示。

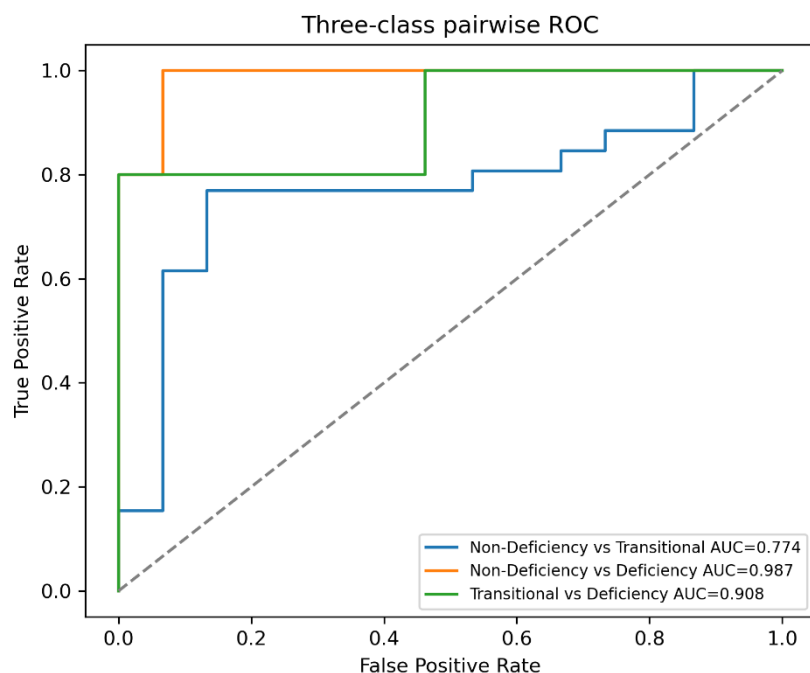


圖-6 獨立測試樣本整體虛證分數之三群 Pairwise ROC 曲線

由圖-6 可知，非虛群與虛群之 AUC 為 0.987，顯示整體虛證分數對兩端群體具有良好之區辨能力。非虛群與過渡群之 AUC 為 0.774，而過渡群與虛群之 AUC 為 0.908，顯示整體虛證分數亦可提供過渡群與相鄰群體之區辨資訊。

本研究於固定整體虛證分數計算公式後，依據訓練樣本之分數分布進行探索性切點分析。分別針對非虛群與過渡群，以及過渡群與虛群進行 pairwise ROC 切點分析，以取得兩群間之建議分數切點，分析結果顯示，非虛群與過渡群之建議切點為 0.698，而過渡群與虛群之建議切點為 0.871。

結論

本研究旨在建立一套結合多模態客觀量測、聚類分析、特徵篩選與整體虛證分數之客觀化分析流程。研究整合肺功能、聲音、行動功能及心率血壓等多模態資料，並採用 FAMvC 進行多模態聚類。初步二群結果顯示，部分受試者之特徵表現介於非虛與虛之間，因此進一步採用三群聚類分析，最終經醫師確認後，分為非虛群、過渡群與虛群。此結果顯示，虛證可能並非單純的二分狀態，而具有由非虛至虛逐漸變化之程度差異。

在特徵篩選方面，本研究採用 Sequential Floating Forward Selection (SFFS)，並透過 10 次重複五折交叉驗證，共取得 50 組特徵子集，以特徵選擇頻率評估其穩定性。比較不同分類器與頻率門檻後，結果顯示 RBF-SVM 於 frequency threshold 為 0.42 時表現最佳，共保留 17 項特徵。

為建立可反映虛證程度之連續量化指標，本研究以非虛群與虛群作為兩端參照，針對篩選後特徵進行 Mann – Whitney U Test，並以 FDR 校正多重比較。結果顯示，PEF、FEV1、Avg. Step Length 與 TUG 等肺功能及行動功能特徵具有顯著差異，聲音與心率血壓特徵亦提供額外資訊。進一步依據特徵差異、方向及權重建構整體虛證分數，分數越高代表整體特徵越接近虛群，反之則越接近非虛群。

將虛證分數套用於獨立測試資料後，三群分數呈現由非虛、過渡至虛逐漸升高之趨勢。Pairwise ROC 分析結果顯示，非虛群與虛群之 AUC 為 0.987，非虛群與過渡群為 0.774，過渡群與虛群為 0.908。依相鄰群體之 ROC 切點，本研究提出探索性建議區間：分數低於 0.698 為較接近非虛狀態，0.698 至 0.871 為較接近過渡狀態，高於 0.871 則較接近虛狀態。

未來可進一步擴大樣本數與資料來源，納入不同醫療場域及族群，以驗證分群結構、關鍵特徵、虛證分數與建議區間之泛化能力；亦可透過縱向追蹤觀察虛證分數隨時間之變化，並整合問診、舌象、脈象等中醫辨證資訊，以建立更完整之多模態虛證客觀化評估架構。