113學年度 元智大學 工業工程與管理學系 畢業專題

運用大語言模型協助醫生評估老人平衡能力：單腳站立的智能輔助系統

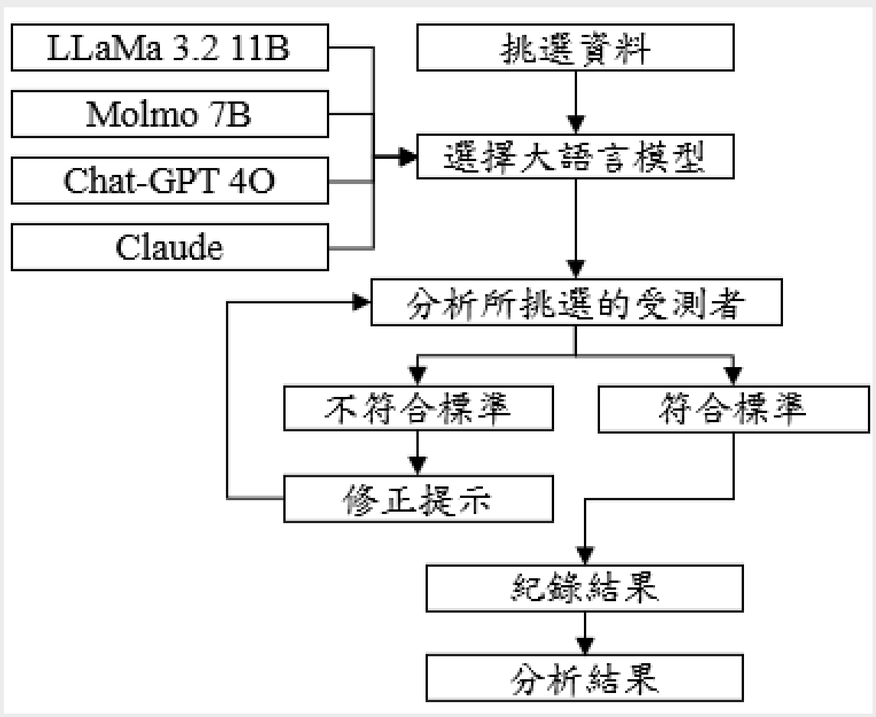
Using Large Language Models to Assist Physicians in Evaluating Elderly Balance Abilities: Intelligent Single-Leg Standing Assessment System

指導教授： 孫天龍 教授　　學生：陳彥均、練錚

研究背景

- 隨著老年化社會發展，老年人跌倒風險增加。
- 傳統人工方法存在主觀性，可能導致評估不一致。

研究架構



評分標準

0分：當受測者無法獨立完成站立，需要外力輔助才能保持平衡或無法將腿抬起時，給予0分。

2分：當受測者無需外力輔助，但站立時間不足10秒，無論姿勢是否穩定，均給予2分。

4分：當受測者能夠獨立站立且時間超過10秒，並且站立過程中無明顯失衡或需要協助，給予4分。

研究動機與目的

動機

- 提升醫療評估效率，降低醫生的工作負擔。
- 利用AI實現更客觀、更準確的分析。

目的

- 開發基於大語言模型的系統，用於單腳站立測試。
- 提升評估準確性，目標準確率超過90%。
- 探討技術應用於臨床決策的可行性。

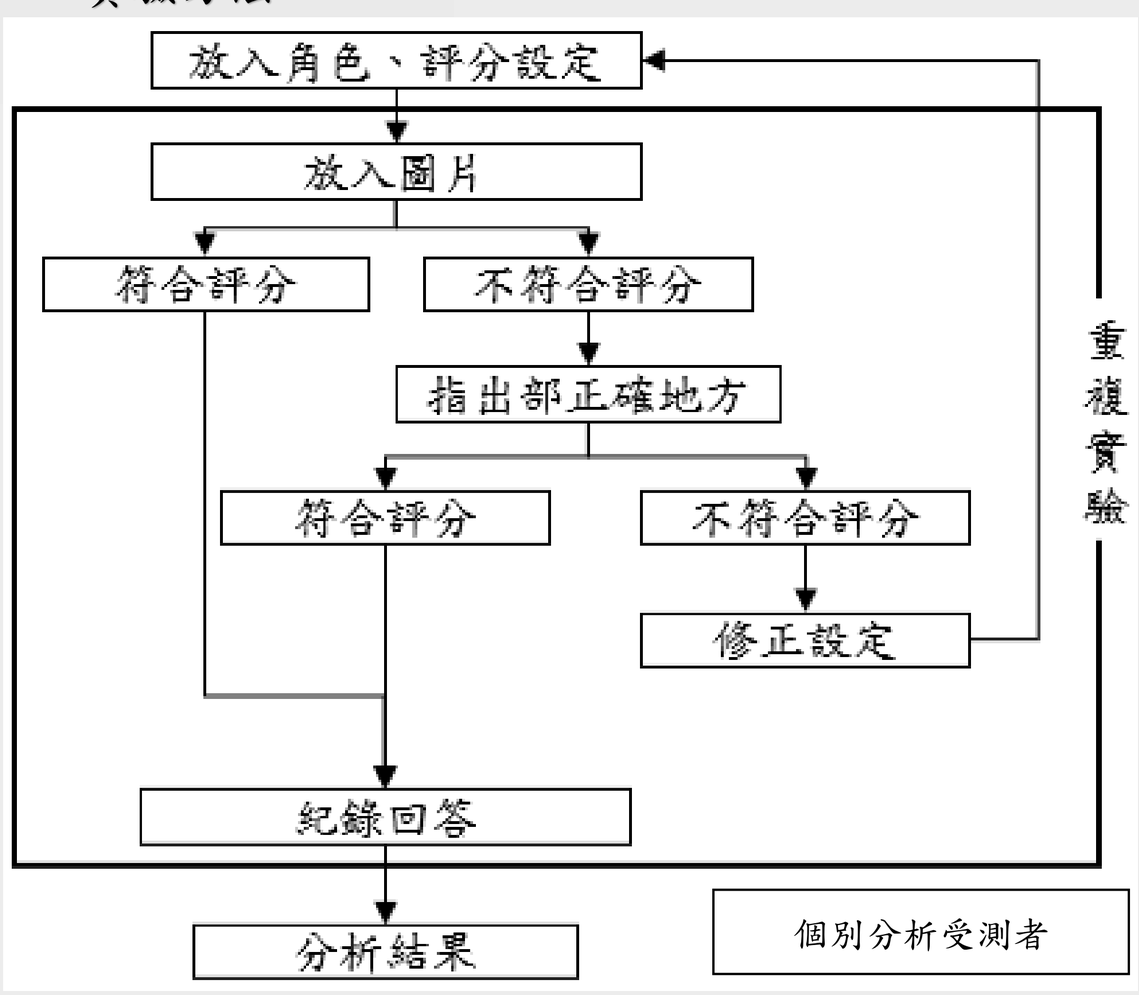
實驗對象及其他設定

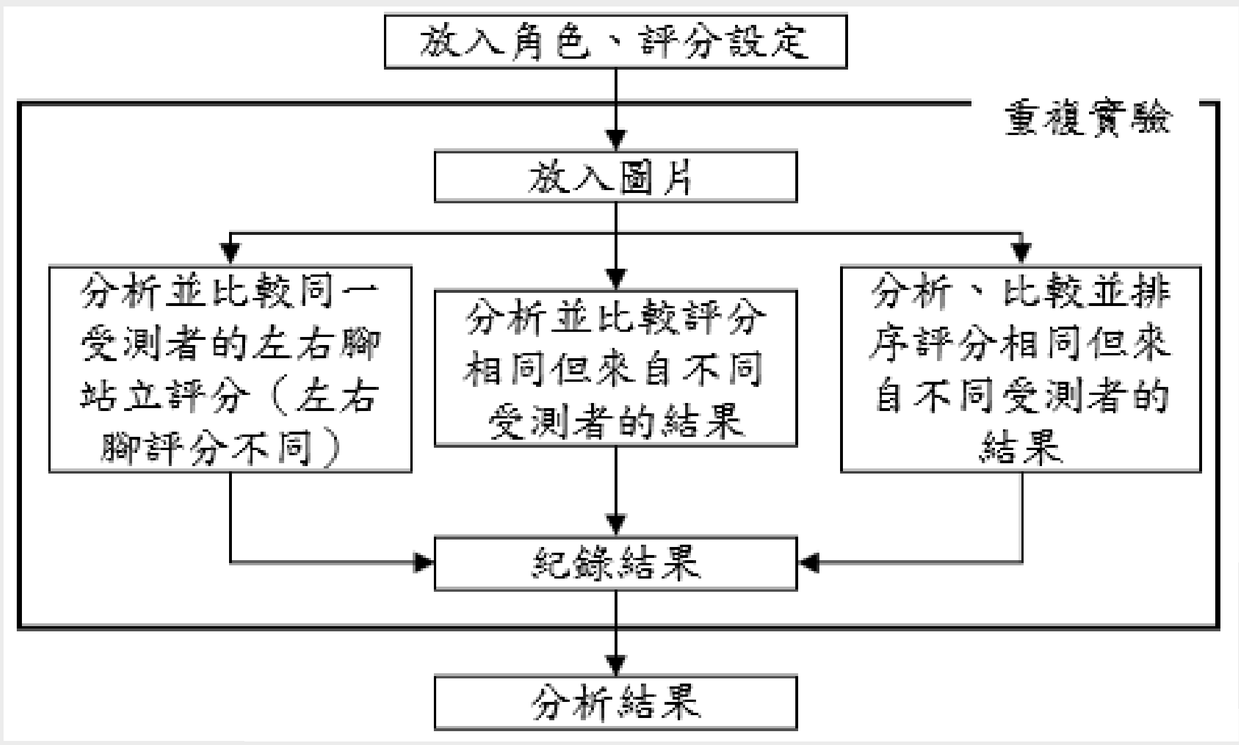
- 分數2分四位
- 分數4分四位
- 三位左右腳分數(例如：左腳評分2分，右腳平分4分)不相同的受測者
- 受測者分數均由醫生先行評分，必以此做為後續判斷大語言模型準確率依據
- 圖片解析度為 1920*1080
- 圖片每0.5秒截一次

實驗模型比較

模型	優點	缺點
ChatGPT	1. 多用途對話和內容生成能力強 2. 支援多語言及多種插件 3. 更新迅速	1. 成本較高 2. 專業領域準確性需調整
Claude	1. 著重內容控制和安全性 2. 對倫理風險控制好，適合敏感領域	1. 生成創意內容稍弱 2. 訓練數據保守，非正式內容生成有限
LLaMA 3.2 11B	1. 開源，易於企業和科研人員自定義 2. 社群參與度高，適合實驗和開發 3. 低資源環境友好	1. 需技術支持、自訂開發 2. 初始穩定性和準確性不如其他商業模型
Molmo 7B	1. 多模態整合能力 2. 開源透明 3. 計算效率高 4. 高質量數據訓練	1. 語言生成能力有限 2. 特定領域泛化不足 3. 高度依賴數據質量

實驗方法





※下表為組內討論後評價。

※ Liama3.2 11B 因為無法產生有效的回答，所以無法評估。

結論

1. 自動化提高效率：AI自動化處理減少了評估所需的時間，並幫助醫生在短時間內得到即時且準確的反饋，提升工作效率。
2. 需要進一步優化：在處理極端情況（如短時間站立或需要外力輔助）時，系統仍有待改善，尤其是在時間判斷和姿勢穩定性分析上。

AI比較	GPT-4o	Claude	Llama3.2 11B	Molmo 7B
成功率	較高	適中	無	低
回答的速度	快	快	適中	適中
語言流暢性	高	高	無	低
回答的豐富度	較差	較好	無	中
一致性(無範本)	低	高	無	低
一致性(有範本)	高	高	無	中