基於EBM可解釋性模型與LLM輔助之臨床AI查房系統

A Clinical Rounding System Empowered by Explainable EBM Models and LLM

指導教授: 蔣榮先教授　合作醫師:劉冠宏醫師、邱裕桓醫師　專題成員: 黃楷軒

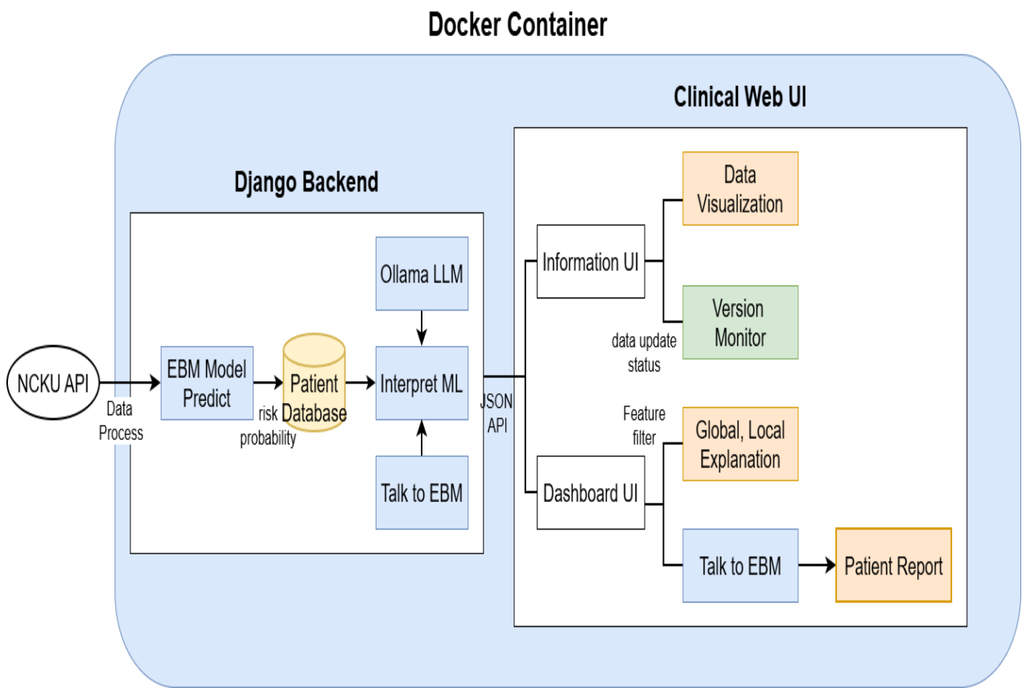
壹. 臨床困境 Clinical Challenges

在透析查房中，醫護人員需快速整合血壓、脫水量、血流速與歷史紀錄，判斷病人是否有透析中低血壓（IDH）風險。然而資料分散、查閱耗時，且缺乏即時可解釋的風險評估工具，容易增加查房負擔並忽略潛在風險。

貳. 研究動機與目的 Research Motivation

專題旨在建立一套以可解釋機器學習模型為核心的 IDH 查房輔助系統，將成大醫院透析 API 的即時資料轉換為臨床可讀的風險資訊。

系統透過 Explainable Boosting Machine預測病人 IDH 風險，並結合interpret ml的特徵貢獻度解釋圖，讓LLM分析摘要報告書狀態，協助醫師快速辨識高風險病人及其可能原因。



圖一: 系統架構圖

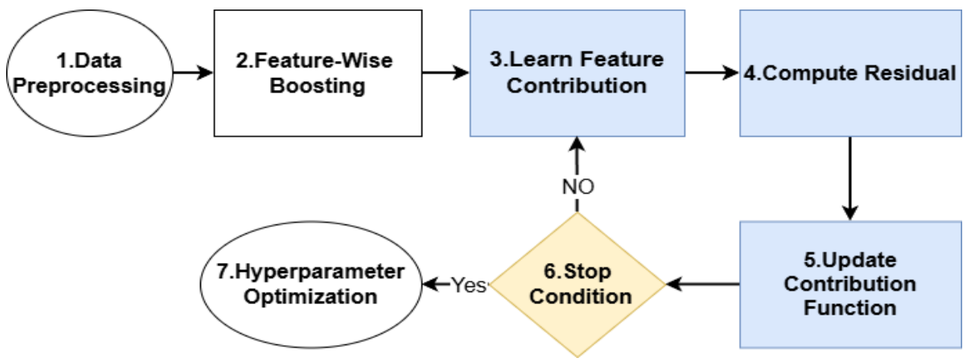
參. EBM模型訓練 EBM Model Training

EBM 模型特色與訓練概念:

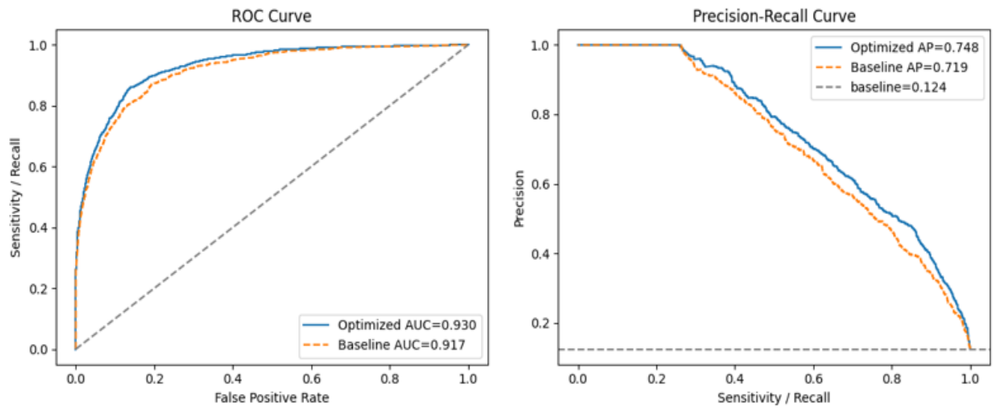
本專題選用 EBM 作為 IDH 風險預測模型，因為它不只輸出風險分數，還能拆解每個臨床特徵對風險的影響，讓醫師理解模型判斷依據。訓練流程會先建立 baseline risk，再將連續特徵分箱，透過 cyclic boosting 逐一更新各特徵函數 $f_i(x_i)$，每輪根據殘差小幅修正模型，最後將所有特徵貢獻加總，轉換成 IDH 發生機率。

訓練調整與優化亮點:

本專題的優化重點不是單純追求準確率，而是降低 IDH 漏判。由於 IDH 事件比例偏低，測試資料中 prevalence 約 12.4%，因此使用 sample_weight 與 pos_weight 提升正類重要性。參數調整時測試 14 組設定，選擇 Recall 最高的模型。最佳化後 Recall 從 41.7% 提升到 74.3%，代表模型更能抓出高風險病人，符合查房輔助中不要漏掉風險的需求。



圖二: EBM訓練架構圖



圖三: ROC 與 Precision-Recall 效能評估曲線

肆. 實作技術與環境 Technical Implementation

後端:

使用 Python、Django 與 REST API，負責資料擷取、前處理、模型推論與 SQL 歷史資料查詢。模型使用EBM，並以 joblib 儲存與載入。

前端:

使用 HTML、CSS、JavaScript 與圖表視覺化呈現病人風險、趨勢與特徵解釋。系統另整合 LLM 、talk to EBM、interpret ML等產生可解釋性呈現，並支援 Docker / Ubuntu 於成大醫院部署與版本監控。

伍. 成果展示 Results and System Demonstration

查房首頁(圖四):

提供病人清單、IDH 風險排序、即時狀態摘要、歷史趨勢折線圖與特徵範圍長條圖，透過風險排序，系統可以優先呈現 IDH 發生機率較高的病人，協助醫師快速辨識高風險病人，更直覺觀察病人近期狀態是否異常。

病人解釋頁(圖五):

利用Interpret ML呈現 EBM 全域與區域解釋，全域解釋(紅框)說明特徵在不同範圍對IDH預測的貢獻度，以及透過平滑曲線和局部最佳解提供調整建議；區域解釋(藍框)則針對單一病人，顯示各特徵如何增加或降低該病人的風險，快速監控危險特徵。系統也支援病人報告書產生功能，利用talk to EBM將全域圖轉成文字形式，結合區域解釋的危險特徵與病人狀態給 LLM gpt-oss模型整理成臨床可讀摘要，大幅增加臨床醫生查房及後續紀錄效率。

圖四: 查房首頁

圖五: 病人解釋頁

陸. 結論與未來展望 Conclusion and Future Work

本系統**將可解釋 AI 應用於透析查房情境，使模型預測不僅提供風險分數，也能轉化為臨床可理解的判讀依據。並且在成大醫院已有部屬和臨床使用案例，定期追蹤與系統維護。**

未來將進一步導入 AI Agent，自動整合病人即時狀態、歷史趨勢與模型解釋，減少醫師在查房前反覆查閱與整理資料的負擔，並協助追蹤異常變化與產生更完整的查房摘要。另一方面，將結合 R Learner 等因果推論方法，針對不同臨床處置與 IDH 風險變化進行分析，補足一般預測模型只能回答「誰高風險」、卻較難回答「採取何種處置可能降低風險」的限制，進一步朝個人化決策輔助發展。

參考文獻：[1] H. Nori et al., "InterpretML: A Unified Framework for Machine Learning Interpretability," arXiv:1909.09223, 2019. [2] Y. Lou, R. Caruana, J. Gehrke and G. Hooker, "Accurate Intelligible Models with Pairwise Interactions," KDD, 2013. [3] R. Caruana et al., "Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission," KDD, 2015.