國立臺北教育大學理學院資訊科學系

國立臺北教育大學  
資訊科學系

專題討論

Department of Computer Science

專題討論

College of Science

National Taipei University of Education

Seminar

可解釋模型與黑盒模型在三高風險預測的實證比較

可解釋模型與黑盒模型在三高風險預測的實證比較

An Empirical Comparison of Interpretable and Black-box Models for Predicting Hypertension, Hyperglycemia, and Dyslipidemia

紀伯喬

Po-Chiao Chi

課程教授：蕭瑛東博士

Advisor: YT Hsiao, Ph. D.

紀伯喬 撰

中華民國 114 年 11 月

114

11

November 2025

國立臺北教育大學教育學院資訊科學系

專題討論

Department of Computer Science

College of Science

National Taipei University of Education

Seminar

可解釋模型與黑盒模型在三高風險預測的實證比較

An Empirical Comparison of Interpretable and Black-box Models for Predicting Hypertension, Hyperglycemia, and Dyslipidemia

紀伯喬

Po-Chiao Chi

課程教授：蕭瑛東博士

Advisor: YT Hsiao, Ph. D.

中華民國 114 年 11 月

November 2025

一、文獻來源

1. 基於血液常規及生化檢測數據的心血管泛疾病機器學習診斷模型構建

Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data

作者： Zhicheng Wang, Ying Gu, Lindan Huang, Shuai Liu, Qun Chen, Yunyun Yang, Guolin Hong, Wanshan Ning

出處： Cardiovascular Diabetology, vol. 23, Article 351, Sep. 2024.

DOI: 10.1186/s12933-024-02439-0

2. 基於Shapley值的可解釋機器學習模型於糖尿病預測之比較研究

A comparative study of explainable machine learning models with Shapley values for diabetes prediction

作者： Kaixin Pang

出處： Healthcare Analytics, vol. 7, 100390, Jun. 2025.

DOI: 10.1016/j.health.2025.100390

3. 應用機器學習預測相對健康成年人第二型糖尿病發生率：台灣10年縱向研究 (Use of machine learning to predict the incidence of type 2 diabetes among relatively healthy adults: A 10-year longitudinal study in Taiwan)

作者： Ying-Qiang Liu, Tzu-Wei Chang, Lung-Chun Lee, Chia-Yu Chen, Pi-Shan Hsu, Yu-Tse Tsan, Chao-Tung Yang, Wei-Min Chu

出處： Diagnostics, vol. 15, no. 1, 72, Jan. 2025.

DOI: 10.3390/diagnostics15010072

二、論文內容摘要

2.1 本文探討的主題

本研究探討如何應用機器學習技術於慢性代謝性疾病（心血管疾病、糖尿病）的風險預測與診斷，並比較可解釋模型與黑箱模型的預測表現，以建立兼具準確性與可解釋性的臨床決策支援系統。

2.2 本文內容摘要

慢性代謝性疾病是全球主要健康負擔，傳統診斷方法成本高且不利於早期篩檢。本研究整合三篇文獻，探討機器學習在疾病預測的應用：

文獻一利用58,616名受試者的血液檢測數據，建立五種機器學習模型診斷69種心血管疾病，XGBoost表現最佳（AUC 0.9921），關鍵預測因子包括鉀、總蛋白、白蛋白等基礎生化指標。

文獻二比較七種可解釋機器學習模型於糖尿病預測，發現MARS（Multivariate Adaptive Regression Splines）結合SHAP分析可提供What-if情境模擬，兼具高準確率與臨床可解釋性。

文獻三針對台灣6,687位健康成年人進行10年縱向追蹤，發現Random Forest和XGBoost在糖尿病預測表現優異，甲狀腺功能為台灣族群的重要預測因子。

三篇研究共同證實機器學習能有效預測慢性代謝性疾病，且可解釋性方法（SHAP、MARS）能增強臨床信任度與實用性。

2.3 本文研究背景

慢性代謝性疾病（包含心血管疾病、糖尿病、高血壓等）是全球首要死因，WHO統計心血管疾病每年造成1,790萬人死亡。傳統診斷方法如心導管檢查、影像學檢查成本高且耗時，無法應用於大規模篩檢。

機器學習技術的發展提供了利用常規血液檢測進行早期診斷的可能性。然而，現有研究存在以下限制：

1. 多聚焦於單一疾病（如僅預測糖尿病或冠心病）
2. 深度學習模型的「黑箱」特性降低臨床信任度
3. 缺乏針對特定族群（如台灣人口）的縱向研究

因此，本研究整合三篇文獻，探討如何建立兼具準確性與可解釋性的機器學習診斷模型。

2.4 本文的假設條件

1. 血液常規及生化檢測數據蘊含足夠的疾病診斷資訊
2. 不同慢性代謝性疾病具有可辨識的血液學特徵差異
3. 可解釋性方法（如SHAP）能有效揭示模型決策邏輯而不損害預測準確率
4. 機器學習模型在不同族群和時間範圍內具有穩定的預測能力

2.5 本文採用的方法

**研究設計**

* 文獻一：回顧性病例對照研究（橫斷面）
* 文獻二：橫斷面研究，比較性分析
* 文獻三：前瞻性隊列研究（縱向追蹤10年）

**研究對象與樣本**

|  |  |  |  |
| --- | --- | --- | --- |
| 文獻 | 樣本數 | 對象 | 追蹤期 |
| 文獻一 | 58,616人 | 中國廈門  （健康25,794人/患者32,822人） | 2018-2023 |
| 文獻二 | 70,692人 | 美國CDC健康調查 | 2010-2018 |
| 文獻三 | 6,687人 | 台灣台中榮總健康成年人 | 2011-2021 |

**特徵變數**

* 血液常規檢測：紅血球、白血球、血小板、血紅素等（22-30項）
* 生化檢測：電解質（鉀、鈉、鈣）、肝腎功能、血脂、血糖等（28-33項）
* 共計50個左右特徵變數

**機器學習演算法**

傳統機器學習：

* Logistic Regression（基線模型）
* Random Forest（集成學習）
* Support Vector Machine
* XGBoost / Gradient Boosting（梯度提升樹）

可解釋性模型：

* MARS（Multivariate Adaptive Regression Splines）
* Genetic Programming

深度學習：

* Deep Neural Network

模型評估指標

* AUC（接收者操作特徵曲線下面積）
* 準確率（Accuracy）
* 靈敏度/召回率（Sensitivity/Recall）
* 特異度（Specificity）
* F1分數

**可解釋性分析**

* SHAP（SHapley Additive exPlanations）：計算每個特徵對預測的邊際貢獻
* MARS What-if分析：模擬特徵值變化對預測結果的影響

2.6 本文使用的資料形式與來源

**資料形式**

結構化電子病歷數據，包含：

* 人口學特徵（年齡、性別）
* 血液常規檢測數值
* 生化檢測數值
* 疾病診斷結果

**資料來源**

|  |  |  |
| --- | --- | --- |
| 文獻 | 資料來源 | 資料特性 |
| 文獻一 | 單一醫學中心（廈門大學附屬第一醫院） | 大樣本、多疾病類型 |
| 文獻二 | 公開資料集（美國CDC健康調查） | 全國代表性樣本 |
| 文獻三 | 單一醫學中心（台中榮總） | 縱向追蹤、台灣族群 |

**資料處理**

* 缺失值處理：移除或插補
* 資料標準化：Z-score或Min-Max正規化
* 資料分割：訓練集、驗證集、測試集（常見比例70:15:15或80:20）

2.7 本文研究結果

**模型表現比較**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 文獻 | 最佳模型 | AUC | 準確率 | 備註 |
| 文獻一 | XGBoost | 0.9921 | 96.18% | 心血管疾病診斷 |
| 文獻二 | MARS+SHAP | - | 高（具體數值未提供） | 糖尿病預測，可解釋性佳 |
| 文獻三 | Random Forest / XGBoost | 0.83-0.85  （推估） | - | 10年糖尿病發生率預測 |

**關鍵預測因子**

**心血管疾病（文獻一）：**

1. 鉀（Potassium）- 維持心臟電生理平衡
2. 總蛋白（Total Protein）- 反映營養狀態
3. 白蛋白（Albumin）- 與心衰相關
4. 間接膽紅素（Indirect Bilirubin）- 抗氧化功能

**糖尿病（文獻二、三）：**

1. 血糖相關指標（空腹血糖、糖化血色素）
2. 血脂指標（總膽固醇、三酸甘油酯）
3. 甲狀腺功能（台灣族群特有發現，文獻三）
4. 肝功能指標

**主要發現**

1. 生化檢測優於血液常規：生化特徵（如電解質、蛋白質）比血液常規（紅白血球計數）更具診斷價值
2. XGBoost表現最佳：在多數任務中，XGBoost優於深度學習和其他傳統機器學習方法
3. 可解釋性與準確率可兼得：MARS+SHAP方法證明可解釋模型不會犧牲預測準確率
4. 族群特異性：台灣研究發現甲狀腺功能為重要預測因子，顯示不同族群需建立本土化模型
5. 早期篩檢可行性：機器學習模型可基於常規血液檢測實現成本效益高的大規模篩檢
6. 多疾病辨識能力：模型不僅能區分健康與疾病，還能辨識不同疾病類型（文獻一證實）

**臨床應用價值**

* 提供低成本、高準確率的早期篩檢工具
* SHAP分析能幫助醫師理解AI決策依據，增強信任度
* What-if分析可模擬治療干預效果
* 適用於大規模人群健康管理

三、閱讀心得

3.1 本文的貢獻

**方法學創新：**

1. 首次系統性比較可解釋模型與黑箱模型於慢性代謝性疾病預測，證實可解釋性與準確率可兼得
2. SHAP與MARS整合應用：文獻二證明MARS結合SHAP可提供What-if情境模擬，突破傳統黑箱模型限制
3. 多疾病診斷系統：文獻一建立涵蓋69種心血管疾病的全面性診斷模型，而非單一疾病預測

**臨床應用價值：**

1. 低成本早期篩檢：證實常規血液檢測結合AI可達高準確率（AUC 0.9921），適合大規模篩檢
2. 關鍵生物標記識別：發現鉀、總蛋白、白蛋白等基礎指標的診斷價值，提醒臨床重視常規檢測
3. 族群特異性發現：文獻三發現甲狀腺功能為台灣族群糖尿病預測的重要因子

**科學貢獻：**

1. XGBoost優越性驗證：跨三篇研究一致證實XGBoost在表格數據上優於深度學習
2. 縱向追蹤證據：文獻三提供10年追蹤數據，證明機器學習可預測未來發病風險
3. 可解釋性突破：解決AI醫療「黑箱」問題，增強臨床醫師信任度

3.2 本文的缺點及優點

優點

研究設計：

* 大樣本研究（58,616人、70,692人、6,687人）提供充分統計效力
* 多演算法系統性比較，客觀選出最佳模型
* 台灣研究提供10年縱向追蹤數據

技術創新：

* 模型可解釋性（SHAP、MARS）解決AI黑箱問題
* What-if分析可模擬治療干預效果
* 卓越診斷準確率（AUC達0.9921）

臨床價值：

* 僅需常規血液檢測，成本低廉易推廣
* 涵蓋多種慢性代謝性疾病（心血管、糖尿病）
* 識別族群特異性危險因子（台灣甲狀腺功能）

缺點

研究設計限制：

* 文獻一、二為橫斷面研究，無法建立因果關係
* 文獻一、三為單中心研究，可能存在選擇偏差
* 缺乏多中心外部驗證，模型泛化能力未知

方法學限制：

* 混雜因子控制不足（生活習慣、用藥史、家族史）
* 文獻一、二缺乏時序性追蹤，無法評估疾病進展
* 未探討模型在不同族群間的表現差異

資料限制：

* 單一時間點測量，無法捕捉生化指標動態變化
* 缺失值處理方法可能影響結果
* 資料不平衡問題（健康人vs患者比例）未充分討論

3.3 本文未來後續發展的方向或空間

驗證性研究

1. 多中心前瞻性驗證：在不同地區、不同醫院、不同族群驗證模型泛化能力
2. 跨國合作研究：比較東西方族群的疾病預測模型差異
3. 臨床隨機對照試驗：評估AI輔助診斷對臨床決策的實際影響

技術改進方向

1. 深度可解釋性研究：
   * 結合注意力機制（Attention Mechanism）與SHAP
   * 開發更直觀的視覺化工具供臨床使用
   * 探討因果推論（Causal Inference）方法
2. 多模態數據整合：
   * 結合影像（CT、MRI、心電圖）
   * 整合基因資訊（多基因風險評分）
   * 納入穿戴裝置連續監測數據
3. 個人化預測模型：
   * 針對不同年齡、性別、種族建立分層模型
   * 開發動態風險評估系統（追蹤生化指標變化）
   * 提供個人化治療建議

臨床應用發展

1. 臨床決策支援系統（CDSS）：整合電子病歷，提供即時風險評估
2. 成本效益分析：評估AI篩檢vs傳統方法的經濟效益
3. 健康管理應用：開發適用於社區、企業健檢的風險分級工具

可解釋性vs準確率權衡研究

1. 探討不同臨床情境的最佳平衡點：急診vs健檢
2. 開發混合模型：結合黑箱模型的準確率與可解釋模型的透明度
3. 醫師接受度研究：評估可解釋性對臨床決策的影響

3.4 如果是你自己，會如何做這個題目

研究設計改進

1. 前瞻性多中心設計：
   * 在台灣北中南東至少5家醫院同時收案
   * 納入城市與鄉村醫療機構，確保樣本代表性
   * 追蹤期至少3-5年，評估預測模型長期效能
2. 更全面的變數收集：
   * 基礎變數：年齡、性別、BMI、血壓、腰圍
   * 生活習慣：吸菸、飲酒、運動、飲食型態
   * 家族史：心血管疾病、糖尿病家族史
   * 用藥史：降壓藥、降血糖藥、降血脂藥
3. 縱向動態監測：
   * 每3-6個月採血追蹤，觀察生化指標變化趨勢
   * 建立時間序列預測模型（LSTM、Transformer）
   * 評估生化指標變化率與疾病發生的關聯

技術方法創新

1. 可解釋性與準確率並重：
   * 主模型：XGBoost + SHAP（兼顧準確率與可解釋性）
   * 輔助模型：MARS提供What-if分析
   * 比較模型：深度學習作為準確率上限參考
2. 模型整合（Ensemble）：
   * 結合多種可解釋模型的優勢
   * 使用Stacking或Blending提升穩健性
   * 針對不同疾病類型選用最佳模型
3. 個人化風險軌跡：
   * 不僅預測「是否發病」，更預測「何時發病」
   * 提供個人化風險曲線
   * 模擬不同干預措施的效果

臨床驗證與應用

1. 嚴格的臨床驗證流程：
   * 內部驗證（Cross-validation）
   * 外部驗證（其他醫院資料）
   * 前瞻性驗證（實際臨床應用）
2. 開發實用工具：
   * 整合電子病歷的CDSS系統
   * 開發手機App供民眾自我評估
   * 提供醫師端與病人端的友善介面
3. 倫理與隱私保護：
   * 去識別化資料處理
   * 符合GDPR與台灣個資法
   * 取得IRB倫理審查核准

3.5 本方法有無其他可以應用或適用的例子

其他慢性疾病預測

|  |  |  |  |
| --- | --- | --- | --- |
| 疾病類型 | 應用方式 | 可行性 | 關鍵檢測項目 |
| 慢性腎病 | 基於肌酸酐、尿素氮等預測腎功能惡化 | 高 | 肌酸酐、eGFR、尿蛋白 |
| 脂肪肝/肝硬化 | 基於肝功能指標預測肝病進展 | 高 | ALT、AST、總膽紅素 |
| 骨質疏鬆 | 結合骨密度與生化指標預測骨折風險 | 中 | 鈣、磷、維生素D |
| 甲狀腺疾病 | 基於TSH、T3、T4預測甲狀腺功能異常 | 高 | TSH、Free T4 |
| 貧血 | 基於血液常規預測貧血類型與原因 | 高 | 血紅素、MCV、鐵蛋白 |

癌症早期篩檢

* 多癌種血液檢測（Multi-cancer early detection）
* 基於腫瘤標記物組合的AI預測
* 肺癌（CEA、NSE）、肝癌（AFP）、攝護腺癌（PSA）

感染性疾病診斷

* 細菌vs病毒感染鑑別（基於白血球分類）
* COVID-19重症風險預測（基於發炎指標）
* 敗血症早期預警（基於多項生化指標）

健康管理應用

1. 企業健檢：員工健康風險分級，提供個人化健康建議
2. 社區篩檢：大規模人群篩檢，識別高危險群
3. 保險風險評估：基於健康數據的保費精算
4. 精準營養：根據生化指標提供飲食建議

藥物反應預測

* 降血壓藥物療效預測
* 降血糖藥物選擇建議
* 抗凝血藥物劑量調整

3.6 其他心得

可解釋性AI的重要性

本研究最大的啟發在於證明可解釋性與準確率並非對立關係。文獻二的MARS+SHAP方法展示了可解釋模型同樣能達到高準確率，且提供What-if分析功能，這對臨床決策至關重要。

傳統觀念認為深度學習準確率高但不可解釋，而可解釋模型（如決策樹）準確率較低。然而，本研究證明XGBoost結合SHAP可以兼得兩者優勢，這對AI醫療的實際應用具有重大意義。

台灣本土化研究的必要性

文獻三的台灣10年縱向研究發現甲狀腺功能為糖尿病預測的重要因子，這在西方研究中較少被強調。這提醒我們：

* 不同族群的疾病危險因子可能不同
* 直接套用國外模型可能失效
* 必須建立台灣本土化的預測模型

AI應作為輔助工具，非取代醫師

儘管研究顯示AI診斷準確率極高，但仍須強調：

1. AI無法取代醫師的臨床判斷：病史詢問、理學檢查、病人溝通無法由AI完成
2. 需要人機協作：AI提供客觀數據分析，醫師提供專業判斷與人文關懷
3. 醫師最終決策權：AI為建議工具，醫師保有最終決策權與責任

未來展望：整合式健康管理

結合本研究的機器學習方法與現代科技（穿戴裝置、遠距醫療），未來可能實現：

1. 連續性健康監測：穿戴裝置追蹤血壓、心率、活動量
2. 定期血液檢測AI分析：每3-6個月健檢數據輸入AI評估風險
3. 個人化健康建議：根據AI分析提供飲食、運動、用藥建議
4. 早期預警系統：風險升高時即時通知，提早介入

倫理與隱私挑戰

AI醫療應用需面對的挑戰：

1. 資料隱私保護：如何在利用大數據訓練模型與保護個人隱私間取得平衡
2. 演算法偏見：避免模型對特定族群產生歧視
3. 責任歸屬：AI誤診時的法律責任如何界定
4. 數位落差：確保偏鄉與弱勢族群也能受惠於AI醫療

個人研究規劃

基於本次文獻閱讀，我未來的研究方向將聚焦於：

1. 可解釋模型與黑箱模型的實證比較：探討在不同臨床情境下的最佳選擇
2. 台灣三高預測本土化模型：利用健保資料庫建立台灣族群專屬模型
3. 多模態整合診斷系統：結合血液檢測、影像、基因資訊的整合性AI系統
4. 臨床實用工具開發：開發真正能在臨床使用的決策支援系統

這三篇文獻為我的研究提供了堅實的理論基礎與方法學參考，證明了機器學習在慢性病預測的巨大潛力，也指出了可解釋性AI的重要性。未來研究將在此基礎上，朝向更精準、更可解釋、更實用的方向發展。

四、參考文獻

慢性代謝性疾病預測（心血管、糖尿病、三高）

*[1] Z. Wang, Y. Gu, L. Huang, S. Liu, Q. Chen, Y. Yang, G. Hong, and W. Ning, "Construction of machine learning diagnostic models for cardiovascular pan-disease based on blood routine and biochemical detection data," Cardiovascular Diabetology, vol. 23, 351, Sep. 2024. doi: 10.1186/s12933-024-02439-0*

*[2] M. Chen, K. Hwang, L. Wang, Y. Hao, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," IEEE Access, vol. 5, pp. 8869–8879, Jun. 2017. doi: 10.1109/ACCESS.2017.2694446*

*[3] J. Yang, X. Ju, F. Liu, O. Asan, T. S. Church, and J. O. Smith, "Prediction for the risk of multiple chronic conditions among working population in the United States with machine learning models," IEEE Open Journal of Engineering in Medicine and Biology, vol. 2, pp. 291–298, Oct. 2021. doi: 10.1109/OJEMB.2021.3117872*

*[4] Y. Luo, Q. Wu, R. Meng, F. Lian, C. Jiang, M. Hu, Y. Wang, and H. Ma, "Associations of serum uric acid with cardiovascular disease risk factors: A retrospective cohort study in Southeastern China," Dryad Dataset, Aug. 2023. doi: 10.5061/dryad.z08kprrk1*

*[5] Y.-Q. Liu, T.-W. Chang, L.-C. Lee, C.-Y. Chen, P.-S. Hsu, Y.-T. Tsan, C.-T. Yang, and W.-M. Chu, "Use of machine learning to predict the incidence of type 2 diabetes among relatively healthy adults: A 10-year longitudinal study in Taiwan," Diagnostics, vol. 15, no. 1, 72, Jan. 2025. doi: 10.3390/diagnostics15010072*

*[6] K. Pang, "A comparative study of explainable machine learning models with Shapley values for diabetes prediction," Healthcare Analytics, vol. 7, 100390, Jun. 2025. doi: 10.1016/j.health.2025.100390*

*[7] M. B. Sampa, M. N. Hossain, M. R. Hoque, R. Islam, F. Yokota, M. Nishikitani, and A. Ahmed, "Blood uric acid prediction with machine learning: Model development and performance comparison," JMIR Medical Informatics, vol. 8, no. 10, e18331, Oct. 2020. doi: 10.2196/18331*

*[8] D. Joedicke, O. Garnica, G. Kronberger, J. M. Colmenar, S. Winkler, J. M. Velasco, S. Contador, and J. I. Hidalgo, "Analysis of the performance of genetic programming on the blood glucose level prediction challenge 2020," in Proc. Genetic and Evolutionary Computation Conference, 2020.*

腎臟疾病預測

*[9] C.-Y. Yang, "運用機器學習演算法和臨床醫療數據以預測腎病症候群患者之病理分類研究 (Using medical data and machine learning algorithms to predict pathologic classification in patients with nephrotic syndrome)," M.S. thesis, School of Medicine, Taipei Medical University, Taipei, Taiwan, Jul. 2023.*

機器學習方法論與應用

*[10] K. Zheng, Z.-H. You, L. Wang, Y.-R. Li, J.-R. Zhou, and H.-T. Zeng, "MISSIM: An incremental learning-based model with applications to the prediction of miRNA-disease association," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 5, pp. 1733–1742, Sep./Oct. 2021. doi: 10.1109/TCBB.2020.3013837*

*[11] A. Kumar, N. Sinha, and A. Bhardwaj, "A novel fitness function in genetic programming for medical data classification," Journal of Biomedical Informatics, vol. 112, 103623, Dec. 2020. doi: 10.1016/j.jbi.2020.103623*

*[12] T.-L. Liou, "疾病預測於保險上之應用:以機器學習的方法建立疾病預測模型 (Application of the disease prediction in insurance: Disease prediction model by machine learning)," M.S. thesis, Dept. Quantitative Finance, National Tsing Hua University, Hsinchu, Taiwan, Jul. 2020.*