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

血清尿酸控制不良增加高血壓發病風險之研究：

中國回顧性隊列研究

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

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

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

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

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

二、論文內容摘要

2.1 本文探討的主題

本研究探討如何利用機器學習技術，基於血液常規和生化檢測數據，建立心血管疾病的診斷模型，以實現早期篩檢和精準診斷。

2.2 本文內容摘要

心血管疾病（CVD）是全球首要死因，傳統診斷方法昂貴且耗時，不利於早期篩檢。本研究利用58,616名受試者的血液檢測數據，建立五種機器學習模型（邏輯迴歸、隨機森林、支援向量機、XGBoost、深度神經網路），用於診斷69種心血管疾病。結果顯示XGBoost模型表現最佳（AUC 0.9921），關鍵預測因子包括鉀、總蛋白、白蛋白和間接膽紅素。研究證實機器學習能有效識別心血管疾病患者，並區分不同疾病類型。

2.3 本文研究背景

心血管疾病是全球最大的健康負擔，尤其在中國。傳統診斷方法如心導管檢查、心電圖、超音波等，成本高且無法有效應用於大規模篩檢。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。過往研究多聚焦於單一疾病風險預測，缺乏心血管疾病的全面性分析。因此，本研究旨在建立能同時診斷多種心血管疾病的機器學習模型。

2.4 本文的假設條件

（本研究無明確假設條件，為觀察性研究）

2.5 本文採用的方法

研究設計：回顧性病例對照研究 資料來源：廈門大學附屬第一醫院（2018-2023年） 研究對象：

健康對照組：25,794人

心血管疾病組：32,822人

涵蓋69種不同心血管疾病

特徵變數：

22項血液常規檢測（如紅血球、白血球、血小板等）

28項生化檢測（如電解質、肝腎功能、血脂等）

共50個特徵變數

機器學習演算法：

邏輯迴歸（Logistic Regression）：基線模型

隨機森林（Random Forest）：集成學習方法

支援向量機（Support Vector Machine）：核函數分類

XGBoost：梯度提升決策樹

深度神經網路（Deep Neural Network）：多層神經網路

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（Sensitivity）

特異度（Specificity）

F1分數

模型解釋：使用SHAP（SHapley Additive exPlanations）算法分析特徵重要性

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，包含血液常規和生化檢測數值

資料來源：單一醫學中心（廈門大學附屬第一醫院）

時間範圍：2018年至2023年

樣本量：58,616人（健康25,794人，患者32,822人）

2.7 本文研究結果

模型表現：

最佳模型：XGBoost

AUC：0.9921

準確率：96.18%

靈敏度：96.90%

特異度：95.26%

關鍵預測因子（依重要性排序）：

鉀（Potassium, K）

總蛋白（Total Protein, TP）

白蛋白（Albumin, ALB）

間接膽紅素（Indirect Bilirubin, NBIL）

主要發現：

生化檢測特徵比血液常規特徵更具診斷價值

模型不僅能區分健康人與患者，還能辨識不同心血管疾病類型

每種心血管疾病具有獨特的血液學特徵

模型提供成本效益高的早期篩檢工具

三、閱讀心得

3.1 本文的貢獻

本研究的主要貢獻在於首次建立基於常規血液檢測的多疾病機器學習診斷系統，涵蓋69種心血管疾病。相較於傳統昂貴的影像學檢查，本研究證明簡單的血液檢測結合AI技術即可達到高準確率（96.18%）診斷。研究提供可解釋的模型（SHAP分析），識別出鉀、總蛋白、白蛋白等關鍵生物標記，為臨床實務提供科學依據。此外，研究比較五種機器學習方法，證實XGBoost在心血管疾病診斷的優越性。

3.2 本文的缺點及優點

優點：

大樣本研究：58,616人提供充分統計效力

多演算法比較：系統性評估五種機器學習方法

高診斷準確率：AUC達0.9921，接近完美

臨床可行性：僅需常規血液檢測，成本低廉

模型可解釋性：SHAP分析提供特徵重要性排序

全面性分析：涵蓋69種心血管疾病，而非單一疾病

缺點：

單中心研究：僅來自一家醫院，可能存在選擇偏差

回顧性設計：無法證明因果關係

缺乏外部驗證：未在其他醫療機構驗證模型泛化能力

時序性不足：橫斷面數據，無法追蹤疾病進展

混雜因子：未完全控制年齡、性別、併發症等變數

3.3 本文未來後續發展的方向或空間

多中心驗證研究：在不同地區醫院驗證模型泛化能力

前瞻性隊列研究：追蹤健康人群，評估模型預測未來發病的能力

整合多模態數據：結合影像、心電圖、基因資訊提升診斷精準度

臨床決策支援系統：開發實用的AI輔助診斷軟體

成本效益分析：評估AI篩檢相較傳統方法的經濟效益

個人化風險預測：根據個人特徵提供精準風險評估

3.4 如果是你自己，會如何做這個題目

如果我進行類似研究，我會考慮以下改進：

前瞻性多中心設計：同時在多家醫院（城市、鄉村）收集數據

縱貫性追蹤：定期採血追蹤，觀察生化指標動態變化

整合臨床變數：納入年齡、性別、BMI、血壓、吸菸史等

深度學習改進：使用Transformer或注意力機制提升準確率

臨床試驗驗證：在真實臨床場景測試模型效能

疾病亞型分析：細分心血管疾病類型，建立專病模型

3.5 本方法有無其他可以應用或適用的例子

本研究的機器學習方法可廣泛應用於其他疾病診斷：

代謝性疾病：糖尿病、高血壓、高血脂的血液檢測診斷

癌症早期篩檢：基於腫瘤標記物的AI預測模型

肝腎功能評估：慢性腎病、肝硬化的風險預測

感染性疾病：透過血液常規識別細菌或病毒感染

神經退化性疾病：阿茲海默症的生物標記預測

健康檢查應用：企業或社區健檢的風險分級

3.6 其他心得

本研究展示人工智慧在醫療診斷領域的巨大潛力，證明常規血液檢測結合機器學習能達到媲美專業影像檢查的診斷準確率。特別值得注意的是，研究發現鉀、白蛋白等基礎生化指標具有極高診斷價值，提醒臨床醫師重視這些常被忽略的指標。 SHAP可解釋性分析是本研究的亮點，解決了深度學習「黑箱」問題，讓AI模型的決策過程透明化，增加臨床醫師的信任度。這對AI醫療的實際應用至關重要。 然而，單中心回顧性研究的侷限性不容忽視。在將模型應用於臨床前，必須進行嚴格的多中心前瞻性驗證。此外，AI模型應作為輔助工具，而非取代醫師判斷。 未來研究方向應朝向整合多模態數據（血液、影像、基因）建立更全面的診斷系統，並開發個人化風險預測模型。結合穿戴裝置的連續監測數據，有望實現心血管疾病的即時預警。

四、參考文獻

[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