國立臺北教育大學理學院資訊科學系

碩士論文(初稿)

Department of Computer Science
College of Science
National Taipei University of Education
Master's Thesis

三高疾病風險預測：多模型比較研究

Risk Prediction of Hypertension, Hyperglycemia, and Dyslipidemia: A Multi-Model Comparative Study

紀伯喬

Po-Chiao Chi

指導教授：許揚博士

Advisor: Yang Syu, Ph. D.

中華民國 115 年 7 月

July 2026

謝辭

脫離學業十年有餘，回想當初考上心目中的臺北科技大學時，抱著由你玩四年的心態，並沒有將重心放在課業上。然而，大學時期工程數學課堂上，楊士萱教授曾說過一句話：「不要對自己太好」，這句話一直銘記在心，卻是出了社會、經歷了職場的磨練後，才真正深刻體悟其中的意義。十多年後，帶著截然不同的心態重返學術殿堂，這一次，我格外珍惜每一堂課、每一次學習的機會。

首先，誠摯感謝指導教授許揚老師的悉心指導。當初選擇軟體工程作為研究方向，正是因為這個領域與我十多年的軟體開發職涯高度契合，這是一個不會後悔的選擇。教授深知在職學生的處境，鼓勵我從工作實務中尋找研究題目，讓學術研究與職場經驗得以相互印證，這樣的指導方式讓我受益良多。

感謝公司大安聯合醫事檢驗所提供的在職進修福利，這份制度成為我重返學術領域的重要推力。公司全額補助學費，大幅降低了經濟上的顧慮，使我能夠專注於學業。同時，順利取得碩士學位後的額外加薪制度，更體現了公司對員工自我提升的重視。我始終相信，所學到的知識是別人帶不走的，而公司投資員工成長、員工回饋所學於工作，正是一個雙贏的局面。

最後，我要特別感謝我的妻子。她是一位獨立且有能力的人，在我每天早出晚歸的求學期間，默默承擔了許多家庭的責任與付出。更令人欣慰的是，在我就讀碩士的第一年，她也報名了國外的碩士進修課程，大兒子進入國小一年級，小兒子進入幼稚園——一家四口同時都是學生，整個家庭充滿了學習的氛圍。這份共同成長的經歷，是這段求學旅程中最珍貴的收穫。

謹以此論文，獻給所有支持我的人。

中文摘要

論文題目：三高疾病風險預測：多模型比較研究

三高疾病（高血壓、高血糖、高血脂）是全球主要的慢性疾病，也是心血管疾病的關鍵可控風險因子。然而，現有風險評估方法多仰賴單一時間點的檢驗數據，未能充分利用縱向健檢資料中蘊含的動態資訊。

本研究使用公開於 Dryad 資料庫的縱向健檢資料集（Luo et al., 2024），涵蓋 6,056 位 40 歲以上社區成人，追蹤期間為 2010 至 2018 年。研究採用三時間點縱貫設計（Y-2、Y-1、Y0），以健檢指標及其變化量特徵（Delta Features, Δ 特徵）預測三高疾病狀態，透過滑動窗口法產生 13,514 筆建模紀錄。本研究系統性比較十種模型（傳統統計、基於實例、樹模型、核方法及神經網路），並以符號回歸 (Symbolic Regression) 探討可解釋性，實驗採用分層分組五折交叉驗證 (StratifiedGroupKFold) 確保無資料洩漏。

主要研究發現：（1）邏輯迴歸 (Logistic Regression) 表現穩定優異，高血糖預測曲線下面積 (Area Under the Curve, AUC) 達 0.938；（2）Δ 特徵可帶來 1.5%–2.3% 的 AUC 提升，且在前十大重要特徵中佔比達 30–50%；（3）SHAP (SHapley Additive exPlanations) 分析揭示疾病特異性預測因子；(4)符號回歸發現極簡公式即可達到 AUC 0.943；（5）累積更多健檢紀錄有助於提升預測性能；（6）僅用前五大特徵，AUC 降幅小於 0.5%。

本研究證實：以縱向健檢資料透過簡單特徵工程與線性模型，即可達到臨床可用的預測性能，適合在基層醫療單位實施早期預警系統。

關鍵詞：三高疾病、高血壓、高血糖、高血脂、機器學習、縱向資料、變化量特徵、SHAP 可解釋性、符號回歸

Abstract

Title: Risk Prediction of Hypertension, Hyperglycemia, and Dyslipidemia: A Multi-Model Comparative Study

Hypertension, hyperglycemia, and dyslipidemia—collectively known as the "three highs"—are major chronic diseases and key modifiable risk factors for cardiovascular disease. Conventional risk assessment methods rely on single-timepoint data, failing to capture the dynamic health trajectories embedded in longitudinal records.

We utilized a publicly available longitudinal dataset (Luo et al., 2024) comprising 6,056 adults aged 40+ followed from 2010 to 2018. A three-timepoint design (Y-2, Y-1, Y0) was adopted, using health indicators and delta features (Δ) to predict disease status. A sliding window approach generated 13,514 records. Eight models (traditional statistics, tree-based, Support Vector Machine (SVM), neural network) were compared, with symbolic regression exploring interpretability. StratifiedGroupKFold 5-fold cross-validation ensured no data leakage.

Key findings: (1) Logistic Regression achieved an Area Under the Curve (AUC) of 0.938 for hyperglycemia; (2) Δ features improved AUC by 1.5%–2.3% and comprised 30–50% of top 10 features; (3) SHapley Additive exPlanations (SHAP) revealed disease-specific predictors; (4) symbolic regression achieved AUC 0.943 with a minimal formula; (5) more checkup records improved performance; (6) top 5 features showed <0.5% AUC decrease.

This study demonstrates that longitudinal health checkup data with simple feature engineering can achieve clinically useful predictions, suitable for early warning systems in primary healthcare.

Keywords: Hypertension, Hyperglycemia, Dyslipidemia, Machine Learning, Longitudinal Data, Delta Features, SHAP Interpretability, Symbolic Regression

目次

表目錄

圖目錄

第一章 緒論

1.1 研究背景

1.1.1 三高疾病的公共衛生重要性

三高疾病，即高血壓（Hypertension）、高血糖（Hyperglycemia）與高血脂（Dyslipidemia），是全球主要的慢性疾病，也是心血管疾病、中風、腎臟病等重大疾病的主要風險因子。根據世界心臟聯盟統計，心血管疾病每年造成約 2,000 萬人死亡，占全球死亡人數近三分之一，其中三高是主要的可控風險因子（World Heart Federation, 2023）。

亞洲地區的三高疾病負擔尤為嚴峻。根據世界衛生組織（World Health Organization, WHO）2023 年全球高血壓報告，西太平洋區域（涵蓋東亞與東南亞）超過四分之一的成年人患有高血壓（WHO, 2023）。在糖尿病方面，中國、日本、印尼等五個亞洲國家的糖尿病患者總數佔全球 48%，且各區域盛行率持續攀升（Ohira & Iso, 2013；JACC: Asia, 2021）。東亞地區的代謝性疾病近數十年急遽增加，其驅動因素包括遺傳易感性、獨特的體脂分布模式，以及都市化與西化飲食的快速轉變（Sun & Zheng, 2025）。值得注意的是，東亞地區的心血管死因以腦中風為主，有別於西亞以缺血性心臟病為主的模式，反映出不同區域三高疾病對心血管系統的影響路徑存在差異（Ohira & Iso, 2013；JACC: Asia, 2021）。

在台灣，三高疾病同樣是重要的公共衛生議題。根據衛生福利部國民健康署 2017-2020 國民營養健康狀況變遷調查，40 歲以上國人高血壓盛行率為 38.3%、高血脂盛行率為 34.1%、高血糖盛行率為 16.4%（國民健康署，2022）。此外，約有 4 至 7 成民眾不知道自己已罹患三高，凸顯早期預測與篩檢的重要性（國民健康署，2022）。

三高疾病往往彼此相關、共同發生，此現象在醫學上被歸納為「代謝症候群（Metabolic Syndrome）」的核心組成。研究顯示，超過 70% 的糖尿病患者同時合併高血壓或高血脂，而糖尿病患者中血脂異常的盛行率更高達 72–85%（Stanciu et al., 2023）。代謝症候群患者的心血管疾病風險為一般人的 2 倍，第二型糖尿病風險則為 5 倍（Alberti et al., 2009）。這種共病現象不僅增加了疾病管理的複雜性，也突顯了同時預測多種疾病風險的重要性。

1.1.2 早期預測與預防的臨床價值

三高疾病在初期往往沒有明顯症狀，患者經常在例行健康檢查或併發症出現時才發現。然而，從健康到發病的過程通常歷經數年的「前驅期」，在此階段透過生活型態調整（如飲食、運動）仍有機會逆轉，研究顯示可降低 50% 以上的發病風險。一旦錯過此窗口，發展為嚴重疾病或併發症後，治療成本將大幅增加。因此，能否在前驅期及早識別高風險個體，是降低三高疾病負擔的關鍵。

三高疾病近年亦呈現明顯的年輕化趨勢。在台灣，30 至 39 歲族群中已有 18.7% 有高血脂、9.7% 有高血壓、2.5% 有高血糖（國民健康署，2022）。為因應此趨勢，國民健康署自 2025 年起將成人預防保健服務年齡從 40 歲下修至 30 歲（國民健康署，2025），反映出早期篩檢與預測的需求已從中高齡族群擴展至青壯年。

1.1.3 縱向健檢資料的研究價值

近年來，隨著健康檢查的普及，大量的縱向健檢資料（Longitudinal Health Checkup Data）被累積。這類資料記錄了同一個體在不同時間點的健康狀態，可追蹤生物標記隨時間的變化趨勢，呈現從健康到疾病的發展過程。由於健康狀態的變化往往先於疾病確診，縱向資料在早期預測上具有獨特的價值。然而，傳統的疾病風險評估方法（如 Framingham 風險評分）主要基於單一時間點的檢驗數據，未能充分利用縱向資料中蘊含的動態資訊。這是本研究欲填補的研究缺口。

1.2 研究目標

1.2.1 主要目標

本研究的主要目標是：

建立一個基於縱向健檢資料的三高疾病預測系統，此系統能夠：

1. **準確預測**：利用歷史健檢資料預測個體未來罹患三高疾病的風險
2. **利用動態資訊**：透過變化量特徵（Δ Features）捕捉健康狀態的動態變化
3. **提供可解釋結果**：識別關鍵風險因子，支持臨床決策

1.2.2 次要目標

在特徵工程方面，本研究驗證變化量特徵（Δ Features）在三高同時預測場景下的有效性，並探索以滑動窗口擴增縱向樣本的可行性。在模型比較方面，系統性比較傳統統計方法、樹模型、核方法、神經網路與符號回歸等多種機器學習模型，評估可解釋性與預測性能之間的權衡。此外，本研究亦驗證多任務學習（MTL）架構同時預測三高疾病的效果，並透過特徵重要性分析識別高風險族群的關鍵特徵，為個人化健康管理提供依據。

1.3 研究貢獻

本研究預期在學術與應用層面做出以下貢獻：

1.3.1 學術貢獻

在縱向特徵工程方面，既有研究（如 Kanegae et al. 2020、Yang et al. 2025）已分別在高血壓與糖尿病預測中證實 Δ 特徵的價值，但尚未有研究將此方法同時應用於三高疾病並進行完整的消融實驗。本研究透過系統性的比較，提供 Δ 特徵在不同疾病間適用性的實證依據。

在模型比較方面，本研究涵蓋傳統統計方法（LR、NB、LDA）、基於實例方法（KNN）、樹模型（DT、RF、XGBoost、LightGBM）、核方法（SVM）、神經網路（MLP）與符號回歸（PySR），此跨類型的系統性比較可為後續研究與臨床應用提供模型選擇的實證依據。本研究亦透過 SHAP 分析與符號回歸探討預測性能與可解釋性之間的權衡，為醫療場景的模型選擇提供指引。

1.3.2 應用貢獻

本研究成果可作為健檢中心部署早期預警系統的基礎，自動標註高風險個體。透過特徵重要性分析所識別的可干預風險因子（如血糖、血壓的變化趨勢），可輔助醫師進行臨床判斷與衛教。此外，符號回歸發現的簡單公式與精簡特徵集，顯示無需複雜模型或昂貴檢驗項目即可達到實用的預測效能，有利於基層醫療單位的實務部署。

1.4 論文架構

本論文共分為八章，各章內容安排如下：

第一章 緒論

說明研究背景、研究目標與預期貢獻，引導讀者了解本研究的定位與價值。

第二章 文獻探討與研究動機

回顧三高疾病預測的相關文獻，涵蓋相關研究總覽、三高疾病預測研究、縱向資料分析與變化量特徵工程，以及類別不平衡處理方法。透過文獻回顧識別現有研究缺口，闡述本研究的動機與定位。

第三章 問題定義

定義本研究欲解決的核心問題，包括現有方法的限制、六個具體研究問題，以及數學化的特徵與預測任務定義。

第四章 研究設計

說明本研究之整體研究架構、資料來源與前處理、特徵工程設計、模型評估指標、實驗設計，以及實驗環境。

第五章 模型方法

詳細說明本研究所採用之各類模型的理論原理與演算法，涵蓋傳統統計方法（LR、NB、LDA）、基於實例方法（KNN）、樹模型（DT、RF、XGBoost、LightGBM）、核方法（SVM）、神經網路（MLP）、符號回歸（PySR），以及類別不平衡處理策略。

第六章 實驗結果

呈現各項實驗的結果，包括模型比較、消融實驗、特徵重要性分析、符號回歸等實驗的數據與觀察。

第七章 討論

針對第六章的實驗結果進行深入討論與詮釋，分析結果的意義、各方法之間的權衡，以及對臨床應用的啟示。

第八章 結論與建議

總結本研究的主要發現與貢獻，討論研究限制，並提出未來研究方向的建議。

第二章 文獻探討與研究動機

本章回顧相關文獻，首先以總覽呈現全貌，接著依序探討三高疾病預測研究、縱向資料與變化量特徵工程，以及類別不平衡處理方法，最後識別研究缺口以闡述本研究動機與定位。各模型理論與實作詳見第五章。

2.1 相關研究總覽

表 2-1 比較本研究與相關文獻的差異

研究	預測目標	資料來源	樣本數	最佳模型	AUC	Δ 特徵	可解釋性
Ye et al. (2018)	高血壓	Maine EHR(美國)	823,627	XGBoost	0.917	無	特徵重要性
Alaa et al. (2019)	心血管疾病	UK Biobank	423,604	AutoPrognosis	0.774	無	—
Dinh et al. (2019)	糖尿病	NHANES	21,131	XGBoost	0.862	無	Info. Gain
Kanegae et al.(2020)	高血壓	日本職場健檢	18,258	XGBoost	0.881	有	特徵重要性
Hung et al. (2021)	隱匿性高血壓	台灣醫院	1,386	RF	0.851	無	—
Liu et al. (2024)	糖尿病	台中榮總 EHR	6,687	XGBoost	0.930	無	特徵重要性
Wang et al. (2024)	高血壓	台灣美兆	207,488	XGBoost	0.889	無	特徵重要性
Yang et al. (2025)	前驅糖尿病	台灣美兆	6,247	XGBoost	—	有	SHAP
Majcherek et al.(2025)	糖尿病	BRFSS (美國)	253,680	Extra Trees	0.99	無	SHAP
本研究	**三高（同時）**	杭州社區調查	6,056	LR / XGBoost	0.721–0.938	**有**	**SHAP + 符號回歸**

註：Ye、Alaa、Kanegae、Liu、Wang、Yang 及本研究為縱向研究設計；Dinh、Hung、Majcherek 為橫斷面研究。

表 2-1 彙整本研究回顧的主要文獻，並以 AUC 作為效能比較基準。由表可觀察到幾項趨勢：（1）XGBoost 在多數研究中達到最佳預測效能；（2）使用 Δ 特徵的研究仍屬少數（僅 Kanegae、Yang 與本研究）；（3）多數研究僅針對單一疾病進行預測，同時涵蓋三高者甚少。各研究的詳細介紹請見後續各節。

2.2 三高疾病預測研究

本研究使用的資料集來自 Luo et al. (2024) 之公開資料，三高疾病的確診狀態依據以下標準標記：高血壓定義為收縮壓（Systolic Blood Pressure, SBP）≥ 140 或舒張壓（Diastolic Blood Pressure, DBP）≥ 90 mmHg，或已確診且正在服用降壓藥物；高血糖定義為空腹血糖（Fasting Blood Glucose, FBG）≥ 7.0 mmol/L 或自我報告糖尿病；高血脂定義為總膽固醇（Total Cholesterol, TC）≥ 6.22 mmol/L。上述閾值與國際通用的診斷標準一致（James et al., 2014；ADA, 2025；NCEP, 2002）。

2.2.1 高血壓預測

Sun et al. (2017) 系統性回顧了 26 篇高血壓預測研究，共涵蓋 48 個預測模型。該回顧指出，常見的風險因子包括身體質量指數（Body Mass Index, BMI）、年齡、血壓水平、吸菸與家族史等，而統計方法以 Logistic Regression（12 篇）、COX Regression（7 篇）和 Weibull Regression（6 篇）為主，顯示傳統統計方法在該領域長期居於主流地位。

近年來，機器學習方法逐漸被引入高血壓預測。Kanegae et al. (2020) 使用日本職場健檢資料（18,258 人）建立高血壓預測模型，採用 XGBoost 和 Ensemble 方法，達到 AUC 0.881。該研究的重要貢獻在於使用縱向變化量特徵（Year(-2) → Year(-1) → Year(0)），證明 Δ 特徵在高血壓預測上的有效性。

Ye et al. (2018) 使用美國 Maine 州的電子健康紀錄（Electronic Health Record, EHR），以 823,627 人的回顧性資料和 680,810 人的前瞻性資料，採用 XGBoost 建立一年期高血壓預測模型，回顧性驗證 AUC 達 0.917，前瞻性驗證 AUC 為 0.870。然而，後續評論指出該研究的前五名重要特徵均為降壓藥物，可能存在資料洩漏問題，提醒研究者在特徵選擇時需審慎避免將結果資訊混入預測因子。

Wang et al. (2024) 使用台灣美兆（MJ）健檢資料進行大規模研究（207,488 人），發現健檢次數越多，預測準確度越高（4 次以上最佳），達到 AUC 0.889。此研究支持多時間點特徵串接的設計理念，與本研究的縱向設計概念一致。

2.2.2 高血糖與糖尿病預測

Liu et al. (2024) 使用台中榮總電子病歷（6,687 人，追蹤 10 年），以 XGBoost 達到 AUC 0.93，關鍵特徵包括糖化血色素（Glycated Hemoglobin, HbA1c）、空腹血糖、體重等。

Yang et al. (2025) 同樣使用 MJ 健檢資料（6,247 位 18-35 歲男性），提出雙框架設計同時預測血糖變化量(δ-FPG)與前驅糖尿病風險。研究發現基線空腹血糖(FPGbase)對預測 δ-FPG 的重要性達 100%，遠超第二名體脂肪的 17.64%，顯示縱向血糖變化具有高度可預測性。本研究的 Δ 特徵設計即參考此概念。

2.2.3 高血脂預測

相較於高血壓與糖尿病，高血脂的機器學習預測研究較少。多數研究將高血脂作為心血管疾病的風險因子，而非獨立的預測目標。本研究將高血脂納入三高同時預測的框架中，填補此研究缺口。

2.3 縱向資料分析與變化量特徵

2.3.1 縱向研究設計

縱向研究（Longitudinal Study）追蹤同一群體在不同時間點的變化，相較於橫斷面研究（Cross-sectional Study）具有以下優勢：

1. **捕捉動態變化**：能觀察生理指標隨時間的趨勢
2. **時序因果關係**：可建立預測因子與結果的時間順序
3. **個體內變異**：控制個體間差異，專注於個體內的變化

然而，縱向資料也面臨挑戰，包括追蹤期間的樣本流失、時間間隔不一致、以及缺失值處理等問題。

2.3.2 變化量特徵工程

變化量特徵（Delta Features）定義為兩個時間點之間生理指標的差值：

$$\delta_j = x_{j,Y_{-1}} - x_{j,Y_{-2}}$$

其中 x_j 為第 j 個生理指標。Yang et al. (2025) 以 δ-FPG（空腹血糖變化量）作為預測目標，證明縱向變化量具有高度可預測性。Kanegae et al. (2020) 同樣使用 Δ 特徵預測高血壓，證明此方法的跨疾病適用性。

本研究採用八個變化量特徵：ΔSBP、ΔDBP、ΔFBG、ΔTC、ΔCr（肌酐酸, Creatinine）、ΔUA（尿酸, Uric Acid）、ΔeGFR（腎絲球過濾率, estimated Glomerular Filtration Rate）、ΔBMI，分別捕捉血壓、血糖、血脂、腎功能與身體質量指數的動態變化。

2.4 類別不平衡處理

三高疾病的發病率通常低於 20%，造成正負類別樣本數量懸殊的類別不平衡（Class Imbalance）問題。在此情境下，模型容易偏向預測多數類（健康），導致少數類（患病）的識別率低落。He & Garcia (2009) 將類別不平衡的處理策略歸納為兩大層面：資料層面與演算法層面。

2.4.1 資料層面方法

資料層面方法透過調整訓練資料的類別分佈來緩解不平衡問題，主要包括過採樣（Over-sampling）與欠採樣（Under-sampling）兩類策略。

過採樣方面，SMOTE（Synthetic Minority Over-sampling Technique）是最具代表性的方法（Chawla et al., 2002）。SMOTE 透過在少數類樣本的特徵空間中進行線性內插，生成合成樣本，避免了簡單複製造成的過擬合問題。其衍生方法包括 Borderline-SMOTE（僅對邊界樣本進行合成）和 ADASYN（根據樣本學習難度自適應生成）。然而，過採樣方法可能引入雜訊樣本，且在高維特徵空間中，合成樣本的品質難以保證。

欠採樣方面，Random Under-sampling 隨機移除多數類樣本以達到類別平衡，但可能丟失重要資訊。Tomek Links 和 Edited Nearest Neighbours（ENN）等方法則透過移除邊界區域的多數類樣本來清理決策邊界，在保留資訊的同時改善類別分離度。

2.4.2 演算法層面方法

演算法層面方法在不改變資料分佈的前提下，透過修改學習演算法本身來處理不平衡問題。

成本敏感學習（Cost-sensitive Learning） 是最常用的演算法層面策略。其核心思想是對不同類別的誤分類賦予不同的代價（cost），使模型更重視少數類的正確分類。在實務上，scikit-learn 等框架提供 class_weight 參數，設定為 'balanced' 時會自動依據類別頻率的倒數調整權重：

$$w_k = \frac{n}{K \cdot n_k}$$

其中 n 為總樣本數，K 為類別數，n_k 為第 k 類的樣本數。此方法的優點在於不改變訓練資料的原始分佈，避免了合成樣本可能引入的雜訊，且計算成本極低。

決策門檻調整（Threshold Moving） 則在模型訓練後，調整分類的機率門檻（預設 0.5）來平衡 Sensitivity 與 Specificity。此方法不影響模型訓練過程，但需要額外的驗證集來選定最佳門檻。

本研究採用 class_weight='balanced' 作為主要的類別不平衡處理策略，並以 AUC-ROC 作為主要評估指標（因 AUC 不受門檻選擇影響），同時報告 Sensitivity 和 Specificity 以反映臨床應用需求。實驗中亦比較了 SMOTE 與 class_weight 兩種策略的效果差異（詳見第六章）。

2.5 研究缺口與本研究定位

2.5.1 相關研究實驗比較

有別於表 2-1 著重預測效能的呈現，表 2-2 從實驗設計的角度檢視各相關研究的涵蓋度，包括系統性比較的模型數量、報告的評估指標、涵蓋的疾病範圍，以及是否進行變化量特徵消融、健檢次數比較、特徵選擇、類別不平衡處理與可解釋性分析等實驗。透過此比較可更清楚地識別現有研究在實驗設計上的不足：多數研究僅比較少量模型且聚焦單一疾病，在特徵選擇與類別不平衡處理等面向的探討亦相當有限。本研究則在上述九個維度中涵蓋全部五項實驗面向，為目前文獻中實驗覆蓋最完整的三高預測研究。

表 2-2 相關研究實驗涵蓋度比較

研究	模型數	評估指標	疾病範圍	Δ 特徵	次數比較	特徵選擇	不平衡處理	解釋性
Ye (2018)	1	4	HTN					✓
Alaa (2019)	9	4	CVD					
Dinh (2019)	5	6	DM					✓
Kanegae (2020)	3	3	HTN	✓				✓
Hung (2021)	4	7	HTN					
Liu (2024)	3	5	DM					✓
Wang (2024)	3	4	HTN		✓			✓
Yang (2025)	5	8	DM	✓				✓
Majcherek (2025)	18	6	DM					✓
本研究	**10**	**4**	**三高**	✓	✓	✓	✓	✓

註：HTN = 高血壓、DM = 糖尿病、CVD = 心血管疾病、三高 = 高血壓 + 高血糖 + 高血脂。「模型數」為研究中系統性比較的模型數量；Auto 表示 AutoPrognosis 自動化模型選擇。「AUC 等四項」指本研究同時報告 AUC、Sensitivity、Specificity 與 F1-Score。

2.5.2 研究缺口

綜觀現有文獻，本研究識別以下研究缺口：

1. **多疾病同時預測**：多數研究僅針對單一疾病，缺乏三高疾病的綜合預測框架
2. **模型比較的完整性**：現有研究通常只比較少數模型，缺乏傳統統計、樹模型、神經網路、符號回歸的全面比較
3. **可解釋性與效能的平衡**：多數研究偏重預測效能，較少探討臨床可解釋性

2.5.3 研究動機

基於上述研究缺口，本研究的動機可歸納為三個面向：

第一，現有三高預測研究多聚焦於單一疾病，且多採用單一時間點資料。然而，三高疾病具有高度共病性（超過 70% 的糖尿病患者合併高血壓或高血脂），且疾病發展為漸進過程，單一時間點的靜態評估無法捕捉健康狀態的動態變化。本研究旨在建立一個同時預測三高疾病的縱向預測框架，透過變化量特徵（Δ Features）充分利用縱向健檢資料的時序資訊。

第二，多數高效能模型（如深度神經網路）屬於黑箱模型，缺乏臨床可解釋性，限制了其在實際醫療場景中的應用。本研究同時探索可解釋方法（符號回歸、SHAP 分析），以期在預測效能與臨床可解釋性之間取得平衡。

第三，隨著台灣將成人預防保健服務年齡下修至 30 歲（國民健康署，2025），早期預測與個人化風險評估的需求日益迫切。本研究期望透過簡單有效的模型與精簡的特徵集，為基層醫療單位提供可實務部署的早期預警方案。

第三章 問題定義

本章定義本研究欲解決的具體問題，包括現有方法的限制、核心與具體研究問題，以及數學化的問題形式定義。

3.1 現有預測方法的限制

目前常見的三高疾病風險評估方法存在以下三大限制：

限制一：單時間點評估

傳統方法僅使用當前的檢驗數據進行風險評估，例如以當前空腹血糖值判斷糖尿病風險。這種方式忽略了一個重要事實：**相同的檢驗數值在不同健康軌跡下具有不同的意義**。例如，血糖從 90 上升到 95 mg/dL 的個體，與從 100 下降到 95 mg/dL 的個體，雖然當前血糖值相同，但其未來風險可能截然不同。

限制二：缺乏縱向資訊

大多數現有研究未充分利用歷史健檢資料，也缺少「變化量特徵」（Delta Features）的工程設計。近期研究（如 Yang et al., 2025）已證實，血糖變化量（δ-FPG）在糖尿病預測中具有極高的重要性，但類似的特徵工程方法尚未被系統性地應用於三高疾病的同時預測。

限制三：模型可解釋性不足

許多高準確度的機器學習模型（如深度神經網路）是「黑箱」模型，難以解釋預測背後的原因。這在醫療應用場景中造成兩個問題：

4. 醫療人員難以信任無法解釋的預測結果
5. 無法提供患者具體可行的風險因子改善建議

3.2 核心研究問題

基於上述背景，本研究提出以下核心研究問題：

如何利用縱向健檢資料，有效預測個體未來罹患三高疾病的風險，同時兼顧預測準確性與模型可解釋性？

3.3 具體研究問題

為回答上述核心問題，本研究設定以下六個具體研究問題：

Q1：變化量特徵的預測價值 在相同的健檢時間點數量下，額外納入變化量特徵（Δ Features）是否能顯著提升三高疾病的預測性能？

Q2：模型選擇與比較 在三高疾病預測任務中，哪些機器學習模型表現最佳？傳統統計方法與機器學習方法的性能差異為何？可解釋模型與黑箱模型之間如何權衡？

Q3：多任務學習的效果 同時預測三高疾病（Multi-Task Learning）是否優於分別預測單一疾病（Single-Task Learning）？

Q4：特徵重要性分析 哪些生物標記及其變化量對三高疾病預測最為重要？這些發現如何支持臨床決策？

Q5：模型精簡的可行性 使用少量關鍵特徵是否能維持接近完整模型的預測性能？此結果對降低臨床資料收集成本有何啟示？

Q6：健檢次數對預測性能的影響 累積更多次健檢紀錄是否能提升預測準確度？此結果對鼓勵民眾定期健檢及健檢機構的服務規劃有何啟示？

3.4 問題形式定義

3.4.1 特徵定義

本研究使用連續三次健檢紀錄，定義以下時間點：

- Y-2：第一次健檢（最早）
- Y-1：第二次健檢
- Y_0：第三次健檢（預測目標時間點）

給定 d 個基本特徵與 p 個健檢指標，輸入特徵向量定義為：

$$X = [X_{base}, X_{Y-2}, X_{Y-1}, \Delta X] \in \mathbb{R}^{d+3p} \tag{3-1}$$

其中：

- $X_{base} \in \mathbb{R}^d$：人口學基本資訊
- $X_{Y-2}, X_{Y-1} \in \mathbb{R}^p$：分別為 Y-2 與 Y-1 時間點的健檢指標
- $\Delta X = X_{Y-1} - X_{Y-2} \in \mathbb{R}^p$：變化量特徵

3.4.2 預測任務

給定輸入特徵 X，學習預測函數 f 使得：

$$\hat{Y} = f(X) = [\hat{y}_{HTN}, \hat{y}_{HG}, \hat{y}_{DL}] \approx Y \in \{0,1\}^3 \quad (3\text{-}2)$$

其中：

- $\hat{y}_{HTN}$：高血壓（Hypertension）
- $\hat{y}_{HG}$：高血糖（Hyperglycemia）
- $\hat{y}_{DL}$：高血脂（Dyslipidemia）

第四章 研究設計

本章說明本研究之整體研究設計，包含研究架構、資料來源與處理、特徵工程、模型評估指標、實驗設計，以及實驗環境。模型方法之詳細說明請參閱第五章。

4.1 研究架構

本研究旨在建立一個基於縱貫性健康檢查資料的三高（高血壓、高血糖、高血脂）風險預測模型。研究架構如圖 4-1 所示，整體流程分為四個階段：資料前處理、特徵工程、模型建立與評估。

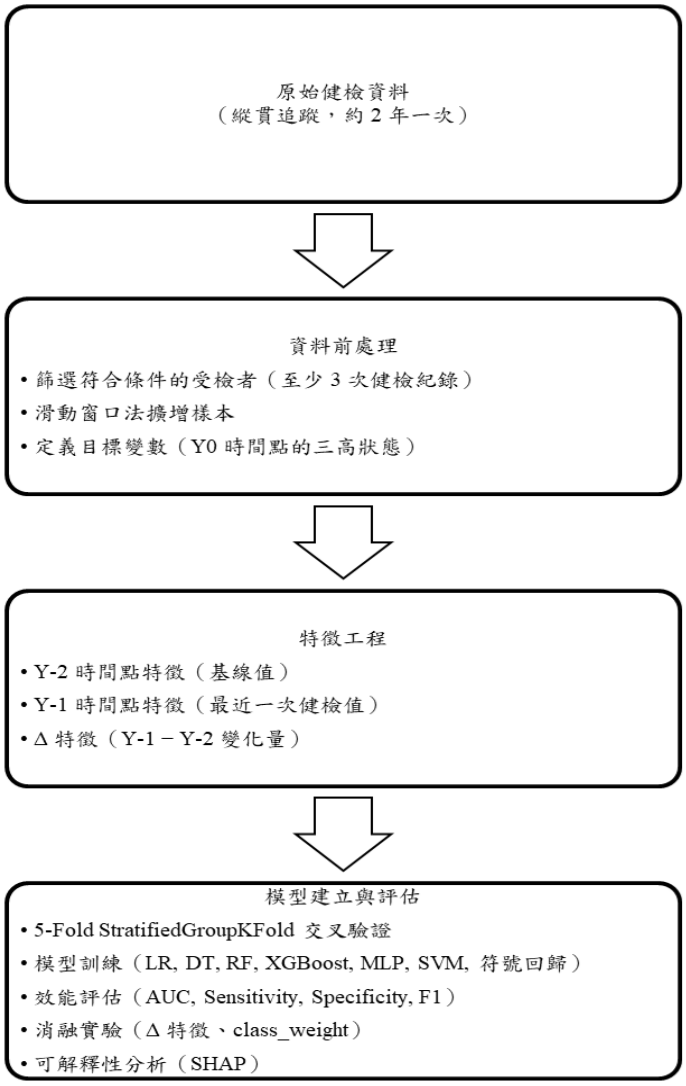

圖 4-1 研究架構圖

4.1.1 研究時間軸設計

本研究採用三個時間點的縱貫設計，如圖 4-2 所示。時間點命名採用相對於預測目標年（Y0）的方式：Y-2 為四年前、Y-1 為兩年前、Y0 為預測目標年。

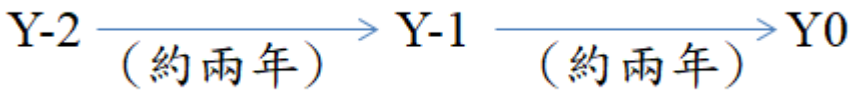

圖 4-2 研究時間軸設計

模型的輸入特徵包含 Y-2 與 Y-1 兩個時間點的健檢指標（SBP、DBP、FBG、TC 等），以及兩時間點之間的變化量（Δ 特徵）。預測目標為 Y0 時間點是否罹患三高（高血壓、高血糖、高血脂）。

選擇 Y0 而非 Y-1 作為預測目標的原因：

1. **避免資料洩漏**：若以 Y-1 為目標，Y-1 的健檢數據與疾病狀態來自同一次檢查，會造成模型「偷看答案」
2. **Δ 特徵可用**：以 Y0 為目標，才能將 Y-1 與 Y-2 的變化量作為有效的預測因子
3. **臨床意義**：提供約 2 年的預警時間窗口，讓醫療人員有足夠時間進行早期介入

4.2 資料來源與處理

4.2.1 資料來源

本研究使用公開於 Dryad 數位資料庫的縱貫性健康檢查資料集（Luo et al., 2024）。

該資料集來自中國浙江省杭州市的社區健康調查，收集期間為 2010 至 2018 年，納入 40 歲以上成人共 6,119 人，多數參與者進行了 3 次以上的健康檢查。

資料特點為僅記錄「第幾次健檢」而無具體日期，追蹤間隔以年齡差推算（例如：55 歲 → 57 歲 = 2 年間隔）。經分析，約 90% 的受檢者維持固定 **2 年間隔**，9.6% 為 1 年間隔（可能為提前回診），平均追蹤間隔為 1.90 年（標準差 0.36 年）。因此，本研究的時間點命名為 Y-2（四年前）、Y-1（兩年前）、Y0（預測目標），反映實際的健檢間隔。由於間隔高度一致，Δ 特徵可直接比較，無需額外的時間校正。

4.2.2 樣本篩選

原始資料集包含 6,119 位參與者共 25,744 筆健檢記錄。由於本研究採用三時間點縱貫設計（Y-2、Y-1、Y0），需要每位參與者至少有 3 次健檢紀錄才能建構完整的特徵集與預測目標。

納入條件： - 至少有 3 次以上的連續健檢紀錄 - 各時間點資料完整，無重大缺失

排除情況： - 共 63 人因僅有 1-2 次健檢紀錄而被排除 - 資料保留率達 98.97%

最終樣本數：6,056 人

篩選後樣本之健檢次數分佈如圖 4-3 所示。約 90% 的樣本健檢次數介於 3 至 5 次之間，其中以 5 次健檢者最多（31.95%），其次為 4 次（29.33%）及 3 次（28.96%）。少數樣本有 6 次以上的健檢紀錄（合計 9.76%）。

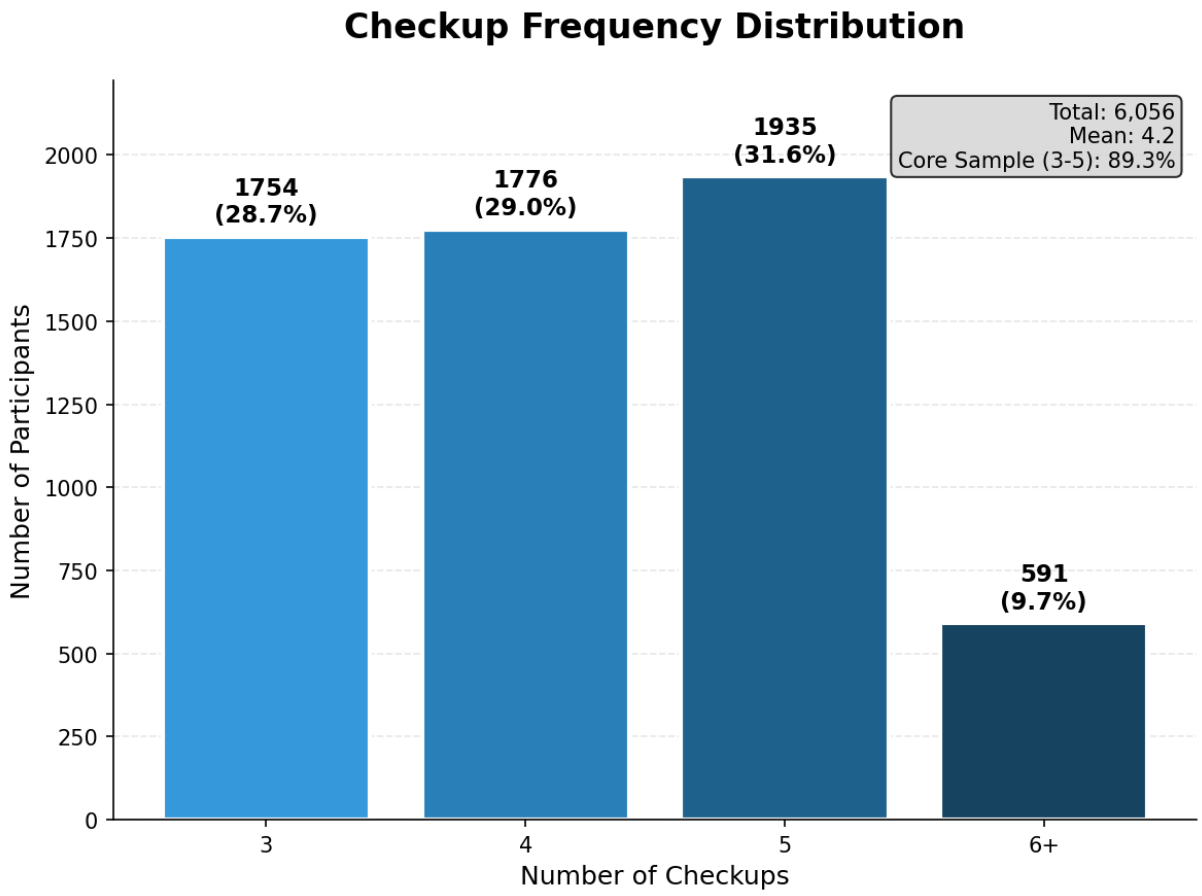

圖 4-3 樣本健檢次數分佈（n = 6,056）

4.2.3 滑動窗口法

為充分利用多次健檢資料，本研究採用滑動窗口（Sliding Window）方法擴增訓練樣本。對於有 N 次健檢紀錄的參與者，可產生 (N-2) 個訓練樣本：

- 3 次健檢 → 1 個樣本：(Y-2, Y-1, Y0)
- 4 次健檢 → 2 個樣本：(Y-3, Y-2, Y-1)、(Y-2, Y-1, Y0)
- 5 次健檢 → 3 個樣本：(Y-4, Y-3, Y-2)、(Y-3, Y-2, Y-1)、(Y-2, Y-1, Y0)

經滑動窗口處理後，6,056 位參與者共產生 **13,514 筆建模紀錄**。此方法的優點：

6. **充分利用資料**：多次健檢者貢獻更多樣本
7. **捕捉不同階段**：同一人在不同年齡階段的健康變化皆納入分析

需注意的是，由於同一參與者可能產生多筆紀錄，在交叉驗證時必須確保同一人的所有紀錄不會同時出現在訓練集與測試集中（詳見 4.4.1 節）。

4.2.4 變數定義

本資料集包含人口學變數、健檢指標及目標變數三類，各變數說明如表 4-1 所示。

表 4-1 研究變數定義

變數類別	變數名稱	說明	單位/編碼
人口學	Sex	性別	1=男, 2=女
	Age	年齡	歲
健檢指標	BMI	身體質量指數	kg/m²
	SBP	收縮壓	mmHg
	DBP	舒張壓	mmHg
	FBG	空腹血糖	mmol/L
	TC	總膽固醇	mmol/L
	Cr	肌酐	μmol/L
	eGFR	腎絲球過濾率	mL/min/1.73m²
	UA	尿酸	μmol/L

本研究之目標變數為三高疾病狀態（高血壓、高血糖、高血脂），由資料集中的確診欄位直接取得。原始資料中，三項目標變數皆以 1 = 正常、2 = 患病 進行編碼，本研究於建模前將其轉換為 0 = 正常、1 = 患病 之二元格式。

4.2.5 類別不平衡情況

三高疾病在本資料集中呈現不同程度的類別不平衡。在 6,056 位樣本中，高血壓患者共 1,010 人（16.68%），負正類比例約為 5:1，屬於輕度不平衡；高血糖患者共 335 人（5.53%），負正類比例約為 17:1；高血脂患者共 361 人（5.96%），負正類比例約為 16:1，兩者皆屬於重度不平衡。

此類別不平衡現象反映了真實世界中三高疾病的盛行率特性，但可能導致模型偏向預測多數類（健康），進而降低對少數類（患病）的識別能力。因此，本研究將於模型訓練階段採用 class_weight 方法進行調整，詳見 3.4.6 節。

4.3 特徵工程

4.3.1 特徵集設計

本研究使用的特徵分為四類，共 26 個特徵，如表 4-2 所示。

表 4-2 特徵集設計

特徵類別	包含特徵	特徵數
基本資訊	Sex, Age	2
Y-2 時間點特徵	FBG_Y-2, TC_Y-2, Cr_Y-2, UA_Y-2, eGFR_Y-2, BMI_Y-2, SBP_Y-2, DBP_Y-2	8
Y-1 時間點特徵	FBG_Y-1, TC_Y-1, Cr_Y-1, UA_Y-1, eGFR_Y-1, BMI_Y-1, SBP_Y-1, DBP_Y-1	8
Δ 特徵（Y-1 − Y-2）	ΔFBG, ΔTC, ΔCr, ΔUA, ΔeGFR, ΔBMI, ΔSBP, ΔDBP	8
合計	—	26

4.3.2 Δ 特徵的意義

Δ 特徵代表 Y-1 與 Y-2 之間的變化量：

$$\Delta_i = X_{i,Y-1} - X_{i,Y-2} \tag{4-1}$$

Δ 特徵的設計理念：

- **捕捉動態趨勢**：某些疾病的發展不僅取決於當前數值，更取決於變化趨勢
- **正值代表上升**：例如 $\Delta\text{FBG} > 0$ 表示血糖在兩年間上升
- **負值代表下降**：例如 $\Delta\text{eGFR} < 0$ 表示腎功能在兩年間下降

4.4 模型評估

4.4.1 交叉驗證策略

本研究採用 scikit-learn 提供的**分層分組 K 折交叉驗證（StratifiedGroupKFold）**進行五折交叉驗證。此方法結合了分層抽樣（Stratified）與群組控制（Group）兩項特性：

1. **分層抽樣**：確保每個 fold 中各類別（患病/健康）的比例與整體資料集一致
2. **群組控制**：確保同一參與者的所有紀錄（由滑動窗口產生）不會同時出現在訓練集與測試集中

此設計的重要性在於：由於滑動窗口法使同一參與者可能貢獻多筆紀錄，若不進行群組控制，模型可能在訓練時學習到某位參與者的特徵模式，而在測試時遇到同一人的其他紀錄，造成評估結果過度樂觀（資料洩漏）。

交叉驗證資料規模

項目	數量
參與者數	6,056 人
建模紀錄數(滑動窗口後)	13,514 筆
Fold 數	5
每 Fold 約測試紀錄數	~2,703 筆

4.4.2 評估指標

AUC-ROC

受試者操作特徵曲線（Receiver Operating Characteristic, ROC）以偽陽性率（False Positive Rate）為 X 軸，真陽性率（True Positive Rate）為 Y 軸，AUC 為曲線下面積。

- AUC = 0.5：隨機猜測
- AUC = 0.7-0.8：可接受
- AUC = 0.8-0.9：良好
- AUC > 0.9：優秀

特點：與分類閾值無關，反映模型的整體排序能力。

敏感度（Sensitivity）

$$Sensitivity = \frac{TP}{TP + FN} \tag{4-2}$$

代表模型正確識別患病者的能力。在疾病篩檢中，高 Sensitivity 意味著較少漏診。

特異度（Specificity）

$$Specificity = \frac{TN}{TN + FP} \tag{4-3}$$

代表模型正確排除健康者的能力。高 Specificity 意味著較少誤診。

F1 分數（F1-Score）

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4-4}$$

精確率（Precision）與召回率（Recall）的調和平均數（Harmonic Mean），適用於類別不平衡的情況。

混淆矩陣

混淆矩陣（Confusion Matrix）為評估分類模型效能的基礎工具，用於呈現模型預測結果與實際類別之間的對應關係。在二元分類問題中，混淆矩陣包含四個元素：

- **True Positive（TP，真陽性）**：實際為患病且模型正確預測為患病的樣本數
- **True Negative（TN，真陰性）**：實際為健康且模型正確預測為健康的樣本數
- **False Positive（FP，偽陽性）**：實際為健康但模型錯誤預測為患病的樣本數
- **False Negative（FN，偽陰性）**：實際為患病但模型錯誤預測為健康的樣本數

上述 Sensitivity、Specificity 及 F1-Score 皆由混淆矩陣計算而得。在健康篩檢應用中，漏診比誤診後果更嚴重，故本研究特別重視 Sensitivity。

未採用 PR-AUC 之說明

在極度類別不平衡（正樣本比例 ＜5%）的情境下，ROC-AUC 可能因大量真陰性而產生過度樂觀的評估，此時 精確率-召回率曲線下面積（Precision-Recall AUC, PR-AUC）被認為是更適合的指標（Saito & Rehmsmeier, 2015）。然而，本資料集三項疾病的盛行率分別為高血壓 16.68%、高血糖 5.53%、高血脂 5.96%，皆高於 5% 的極度不平衡門檻，且本研究已搭配 class_weight 調整與 Sensitivity/Specificity 報告，足以反映模型在少數類上的辨識能力，故本研究以 AUC-ROC 作為主要評估指標。

4.5 實驗設計

本研究設計一系列消融實驗（Ablation Study）與比較實驗，以驗證研究問題。實驗設計總覽如表 4-3 所示。

表 4-3 實驗設計總覽

實驗	目的
Δ 特徵消融	驗證變化量特徵的貢獻
模型比較	比較簡單與複雜模型效能
特徵選擇消融	驗證精簡特徵集的可行性
類別不平衡處理比較	比較不同不平衡處理方法
符號回歸實驗	探索可解釋數學公式

4.5.1 消融實驗

Δ 特徵消融實驗

為驗證 Δ 特徵對預測效能的貢獻，本研究設計五組特徵組合進行消融實驗。

表 4-4 Δ 特徵消融實驗設計

實驗組	特徵組合	特徵數	說明
Full	Y-2 + Y-1 + Δ	26	完整特徵集
No-Δ	Y-2 + Y-1	18	移除 Δ 特徵
Y-2-Only	Y-2	10	僅使用基線值
Y-1-Only	Y-1	10	僅使用最近值
Δ-Only	Δ	10	僅使用變化量

特徵選擇消融實驗

為驗證精簡特徵集的可行性，基於 SHAP 重要性排序設計消融實驗。

表 4-5 特徵選擇消融實驗設計

實驗組	特徵數	說明
Top 3	3	僅使用最重要的 3 個特徵
Top 5	5	僅使用最重要的 5 個特徵
Top 10	10	使用前 10 個重要特徵
All	26	使用全部特徵（基準線）

各疾病的 Top 5 特徵由 SHAP 分析結果決定，詳見第六章。

4.5.2 類別不平衡處理比較

由於三高疾病的患病率較低（5-17%），模型容易偏向預測多數類（健康），導致 Sensitivity 偏低。本研究比較五種類別不平衡處理方法，如表 3-6 所示。

表 4-6 類別不平衡處理方法比較

方法	類型	原理	特點
Baseline	無處理	使用原始資料分佈	作為對照基準
class_weight	權重調整	調高少數類損失函數權重	不改變資料分佈
SMOTE	過採樣	特徵空間中合成少數類樣本	增加訓練樣本數
ADASYN	自適應過採樣	針對邊界少數類合成新樣本	關注邊界樣本
RandomUnderSampler	欠採樣	隨機移除多數類樣本	可能損失資訊

class_weight 權重計算：

$$w_i = \frac{n_{samples}}{n_{classes} \times n_{samples_i}} \tag{4-5}$$

其中 w_i 為第 i 類的權重，$n_{samples}$ 為總樣本數，$n_{classes}$ 為類別數，$n_{samples_i}$ 為第 i 類的樣本數。

SMOTE 演算法原理：

1. 對每個少數類樣本，找出其 k 個最近鄰（預設 k=5）
2. 隨機選擇一個最近鄰
3. 在原樣本與選定鄰居之間的連線上隨機生成新樣本
4. 重複直到少數類與多數類樣本數平衡

4.5.3 符號回歸實驗

符號回歸旨在從資料中自動發現可解釋的數學公式。本研究使用 PySR 套件進行符號回歸實驗，實驗設計如表 4-7 所示。

表 4-7 符號回歸實驗設計

參數	設定	說明
套件	PySR	基於 Julia 的符號回歸套件
二元運算子	+, -, *, /	基本四則運算
一元運算子	exp, log, abs, square	數學轉換函數
最大複雜度 (maxsize)	35	控制公式長度上限
迭代次數 (niterations)	200	遺傳演算法迭代次數
複雜度懲罰 (parsimony)	0.0001	避免過於簡單的常數解
族群數 (populations)	20	平行搜索的族群數量
族群大小 (population_size)	100	每個族群的個體數

實驗流程：

1. 使用 5-Fold StratifiedGroupKFold 交叉驗證（按 patient_id 分組）
2. 對訓練集進行標準化（StandardScaler）
3. 在每個 fold 的訓練集上執行 PySR
4. 從 Pareto 前沿選擇最佳公式（平衡複雜度與準確度）
5. 將預測值限制在 [0, 1] 區間作為機率估計
6. 使用訓練集正樣本比例作為分類閾值
7. 在測試集上評估公式的 AUC

公式評估標準：

- **預測效能**：AUC 與 Logistic Regression 相近（差距 < 5%）
- **穩定性**：多個 fold 產出類似的公式結構
- **可解釋性**：公式符合臨床直覺（如：SBP ↑ → 高血壓風險 ↑）

4.5.4 可解釋性分析

本研究使用 SHAP (SHapley Additive exPlanations) 進行模型可解釋性分析：

- **SHAP 值**：量化每個特徵對預測結果的貢獻
- **特徵重要性排序**：識別最具影響力的風險因子
- **交互效應**：分析特徵間的協同或拮抗作用

4.6 實驗環境

本研究之實驗環境與使用套件如表 4-8 所示。

表 4-8 實驗環境與工具

類別	項目	規格/版本
硬體環境	處理器	Intel Core i7-11700 @ 2.50GHz (8 核心)
	記憶體	32 GB DDR4 3200 MHz
	顯示卡	NVIDIA GeForce RTX 3050 (6 GB VRAM)
	儲存裝置	SSD (ADATA SX8200PNP + WDC WDS200T2B0A)
軟體環境	作業系統	Windows 10 專業版
	程式語言	Python 3.10
	開發環境	Jupyter Notebook, VS Code
主要套件	機器學習	scikit-learn, XGBoost, LightGBM
	神經網路	MLPClassifier (scikit-learn)
	符號回歸	PySR（初期曾使用 gplearn）
	可解釋性	SHAP
	資料處理	pandas, numpy
	視覺化	matplotlib, seaborn

第五章 模型方法

本研究採用十一種預測模型，依方法論性質分為三大類別：傳統統計方法、機器學習方法，以及符號回歸。分類架構如下：

- **傳統統計方法**：Logistic Regression (LR)、Naive Bayes (NB)、Linear Discriminant Analysis (LDA)
- **機器學習方法**
    - 基於實例：K-Nearest Neighbors (KNN)
    - 樹狀模型：Decision Tree (DT)、Random Forest (RF)、XGBoost、LightGBM
    - 核方法：Support Vector Machine (SVM)
    - 神經網路：Multi-Layer Perceptron (MLP)
- **符號回歸**：PySR

傳統統計方法具有明確的數學假設與高度可解釋性，適合作為基準模型。機器學習方法從樹狀模型的離散分割、核方法的連續決策邊界，到神經網路的多層非線性轉換，逐步提升模型的表達能力。符號回歸則嘗試從資料中演化出可解釋的數學公式，兼顧預測能力與可解釋性。以下依序介紹各類模型的原理、選用原因與實作設定。

5.1 傳統統計方法

傳統統計方法具有明確的數學形式與統計假設，模型參數可直接解讀其物理意義。本研究選用三種具代表性的線性分類方法：Logistic Regression 為判別式模型，直接建模後驗機率；Naive Bayes 為生成式模型，透過貝氏定理估計類別機率；Linear Discriminant Analysis 則透過最大化類別間距進行分類。三者在理論上形成互補的比較基礎。

5.1.1 Logistic Regression (LR)

Logistic Regression（邏輯斯迴歸）是疾病預測研究中最常用的基準模型，屬於判別式模型（Discriminative Model），直接建模特徵與類別之間的條件機率。根據 Sun et al. (2017) 的系統性回顧，在 26 篇高血壓預測研究所涵蓋的 48 個模型中，Logistic Regression 佔 12 篇（25%）為最大宗，顯示其在疾病風險預測領域的主流地位。

模型形式：

LR 透過 sigmoid 函數將線性組合映射至 [0, 1] 區間，輸出疾病發生的機率：

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}} \tag{5-1}$$

其中 β_i 為迴歸係數，e^{β_i} 可直接解釋為勝算比（Odds Ratio），表示第 i 個特徵每增加一個單位時，疾病風險的倍數變化。此特性使 LR 在臨床應用中具有高度可解釋性，醫療人員可直觀理解各風險因子的貢獻程度。

在 Sun et al. (2017) 回顧的研究中，LR 的 C-statistic（等同 AUC）多落在 0.72–0.85 之間，顯示即使在非線性關係存在的情境下，LR 仍能提供具競爭力的預測效能。然而，LR 假設特徵與對數勝算之間為線性關係，可能無法捕捉複雜的非線性交互作用。

選用原因：

- 可解釋性高：迴歸係數可直接解讀為風險因子的貢獻（Odds Ratio）
- 計算效率高：適合作為基準模型
- 支援 class_weight：可處理類別不平衡問題

實作參數：

- **solver**: 'lbfgs'
- **max_iter**: 1000
- **class_weight**: 'balanced'（處理類別不平衡）

5.1.2 Naive Bayes (NB)

Naive Bayes（單純貝氏分類器）是一種基於貝氏定理的生成式模型（Generative Model），透過估計各類別下特徵的聯合分佈進行分類。與 Logistic Regression 的判別式模型形成理論上的互補，兩者的比較有助於理解資料的分佈特性。

演算法原理：

NB 基於貝氏定理進行分類：

$$P(Y = k|X) = \frac{P(X|Y = k) \cdot P(Y = k)}{P(X)} \tag{5-2}$$

其核心假設為各特徵在給定類別下條件獨立，即：

$$P(X|Y=k) = \prod_{j=1}^{n} P\left(X_j|Y=k\right) \tag{5-3}$$

本研究使用 Gaussian Naive Bayes，假設連續特徵在各類別下服從常態分佈：

$$P(X_j|Y=k) = \frac{1}{\sqrt{2\pi\sigma_{jk}^2}} \exp\left(-\frac{(X_j-\mu_{jk})^2}{2\sigma_{jk}^2}\right) \tag{5-4}$$

此方法的計算效率極高（時間複雜度為 $O(nd)$，n 為樣本數、d 為特徵數），且無需迭代優化，在小樣本情境下表現穩健。然而，特徵獨立假設在醫療資料中往往不成立，例如收縮壓與舒張壓、血糖與 BMI 之間均存在相關性，此假設的違反可能影響機率估計的校準度，但對分類排序（AUC）的影響通常較小。

選用原因：

- 計算效率極高：僅需估計各特徵的均值與變異數
- 理論基礎明確：基於機率推論框架
- 適合作為基準：與 LR 的判別式模型形成互補比較

實作參數：

- 使用 scikit-learn 的 **GaussianNB**
- **priors**: None（依訓練資料自動估計類別先驗機率）

注意：Naive Bayes 的條件獨立假設在實務中通常不完全成立（如 SBP 與 DBP 高度相關），但即便假設違反，其分類表現仍可作為有意義的參考基準。

5.1.3 Linear Discriminant Analysis (LDA)

線性判別分析（Linear Discriminant Analysis, LDA）由 Fisher (1936) 提出，是統計學中最經典的分類方法之一。LDA 屬於生成式模型，透過最大化類別間變異與類別內變異的比值，尋找最佳線性投影方向進行分類。

演算法原理：

LDA 尋找投影方向 w，使得 Fisher 準則最大化：

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{5-5}$$

其中 S_B 為類別間散布矩陣（Between-class scatter matrix），S_W 為類別內散布矩陣（Within-class scatter matrix），w 為投影方向。LDA 假設各類別的特徵服從多變量常態分佈且共享相同的共變異數矩陣。

相較於 Logistic Regression，LDA 同時考慮特徵的聯合分佈結構，在特徵間存在多重共線性時仍能維持穩定性。此外，LDA 的降維特性（將 d 維特徵投影至最多 $k-1$ 維空間，k 為類別數）使其在高維度資料中具有正則化效果。然而，常態分佈與等共變異數假設在實際資料中可能不完全成立，限制了其對非線性關係的捕捉能力。

選用原因：

- 兼具降維與分類功能：可同時降低特徵維度
- 考慮類別分佈結構：利用共變異數矩陣進行判別
- 計算效率高：無需迭代優化

實作參數：

- 使用 scikit-learn 的 **LinearDiscriminantAnalysis**
- **solver**: ‘svd’（奇異值分解，適合特徵數多於樣本數的情況）
- **priors**: None（依訓練資料自動估計）

5.2 機器學習方法

相較於傳統統計方法的線性假設，機器學習方法能捕捉更複雜的非線性關係與特徵交互作用。本節依序介紹基於實例的方法（KNN）、樹狀模型（DT、RF、XGBoost、LightGBM）、核方法（SVM）與神經網路（MLP），四類方法在學習策略與模型複雜度上各有特色：基於實例的方法不建立顯式模型，直接利用訓練資料進行預測；樹狀模型透過離散分割建立規則；核方法在高維空間中建立連續決策邊界；神經網路則透過多層非線性轉換學習抽象的特徵表示。

5.2.1 K-Nearest Neighbors (KNN)

K-最近鄰（K-Nearest Neighbors, KNN）是一種基於實例的學習方法（Instance-based Learning），屬於惰性學習（Lazy Learning）——訓練階段不建立顯式模型，而是在預測時直接從訓練資料中搜尋最相似的樣本進行分類。

演算法原理：

給定一個待分類樣本 x，KNN 在訓練集中找出距離 x 最近的 k 個鄰居，以這 k 個鄰居的多數類別作為預測結果：

$$\hat{y} = \underset{c}{\text{argmax}} \sum_{i \in N_k(x)} \mathbb{1}(y_i = c) \tag{5-6}$$

其中 $N_k(x)$ 為 x 的 k 個最近鄰集合，$\mathbb{1}$ 為指示函數。距離度量採用歐氏距離（Euclidean Distance）：

$$d(x, x') = \sqrt{\sum_{j=1}^{d} (x_j - x'_j)^2} \tag{5-7}$$

KNN 的預測品質高度依賴 k 值的選擇與特徵的尺度。k 值過小容易受雜訊影響，過大則可能模糊類別邊界。由於使用歐氏距離，不同量綱的特徵需進行標準化處理。

選用原因：

- 學習範式獨特：不建立參數模型，與其他方法形成對比
- 概念直觀：「相似的人有相似的健康風險」符合醫學直覺
- 實作簡便：scikit-learn 提供 **KNeighborsClassifier**

實作參數：

- **n_neighbors**: 5
- **metric**: ‘minkowski’（p=2，即歐氏距離）
- **weights**: ‘uniform’（等權投票）

注意:KNN 不直接支援 class_weight 參數。本研究在預測階段使用距離加權投票，並搭配標準化特徵以確保各維度的距離貢獻均等。

5.2.2 Decision Tree (DT)

決策樹是一種基於規則的分類模型，透過遞迴地將資料依特徵值分割成子集，最終形成樹狀結構，每個內部節點代表一個特徵的判斷條件，每個葉節點代表一個類別預測。

演算法原理：

決策樹的建構過程為：在每個節點選擇使資料純度最大化的特徵與閾值進行分割。本研究使用 Gini Impurity（基尼不純度）作為分割準則：

$$Gini(t) = 1 - \sum_{k=1}^{K} p_k^2 \tag{5-8}$$

其中 p_k 為節點 t 中第 k 類樣本的比例。Gini 值越小代表節點越純，當節點中所有樣本屬於同一類別時，$Gini = 0$。在每次分割時，演算法選擇使加權 Gini 值下降最多的特徵與閾值：

$$\Delta Gini = Gini(parent) - \frac{n_{left}}{n} Gini(left) - \frac{n_{right}}{n} Gini(right) \tag{5-9}$$

分割過程遞迴執行，直到滿足停止條件（如深度限制或樣本數不足）。

選用原因：

- 高度可解釋：分類規則可直接呈現為 if-then 規則
- 計算效率高：訓練與預測速度快
- 支援 class_weight：可處理類別不平衡問題

實作參數：

- **criterion**: ‘gini’
- **max_depth**: None（完全生長）
- **class_weight**: ‘balanced’

注意：單一決策樹容易過擬合，預測效能通常低於集成方法，但因其高可解釋性，仍納入比較。

5.2.3 Random Forest (RF)

Random Forest（隨機森林）是由 Breiman (2001) 提出的集成學習方法，透過結合多棵決策樹的預測來提升效能與穩定性。其核心思想為：多個弱學習器（個別決策樹）透過集成可形成強學習器。

演算法原理：

Random Forest 基於 Bagging（Bootstrap Aggregating）策略：

1. **Bootstrap 抽樣**：從原始資料中有放回地抽樣產生 B 個子資料集
2. **隨機特徵選取**：在每棵樹的每次分裂時，僅從隨機選取的 m 個特徵（$m \ll d$，d 為總特徵數）中選擇最佳分割
3. **集成預測**：最終預測為所有樹的多數決（分類）：

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \ldots, h_B(x)\} \tag{5-10}$$

其中 $h_b(x)$ 為第 b 棵決策樹的預測結果。Bootstrap 抽樣引入的隨機性降低了模型的變異數，而隨機特徵選取進一步降低了樹之間的相關性，使集成效果更為顯著。

選用原因：

- 抗過擬合：Bagging 機制降低變異數
- 穩定性高：對異常值和雜訊較不敏感
- 支援 class_weight：可處理類別不平衡問題

實作參數：

- **n_estimators**: 100
- **max_depth**: None（完全生長）
- **class_weight**: ‘balanced’

5.2.4 XGBoost

XGBoost（eXtreme Gradient Boosting）由 Chen & Guestrin (2016) 提出，是一種基於梯度提升（Gradient Boosting）的集成學習方法。與 Random Forest 的平行集成（Bagging）不同，XGBoost 採用序列式集成策略，每棵新樹專注於修正前面樹的預測誤差，在多項醫學預測任務中表現優異。

演算法原理：

梯度提升透過逐步加入決策樹，累積修正預測結果：

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{5-11}$$

其中 f_t 為第 t 棵新加入的決策樹。XGBoost 的目標函數同時考慮預測誤差與模型複雜度：

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega\left(f_k\right) \tag{5-12}$$

其中 l 為損失函數，$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2$ 為正則化項（T 為葉節點數，w 為葉節點權重），用於控制模型複雜度、防止過擬合。

選用原因：

- 預測效能強：在許多醫學預測任務中表現優異
- 可處理非線性關係：能捕捉特徵間的複雜交互作用
- 內建正則化：目標函數含正則化項，有效防止過擬合
- 支援特徵重要性評估

實作參數：

- **n_estimators**: 100
- **max_depth**: 6
- **learning_rate**: 0.1
- **scale_pos_weight**: 自動計算（處理類別不平衡）

5.2.5 LightGBM

LightGBM（Light Gradient Boosting Machine）由 Ke et al. (2017) 提出，同屬梯度提升樹家族，但針對訓練效率進行了兩項關鍵優化，使其在大規模資料集上的訓練速度顯著快於 XGBoost，同時維持相近的預測效能。

演算法原理：

LightGBM 的梯度提升框架與 XGBoost 相同，但在樹的建構策略上引入兩項加速技術：

1. **基於梯度的單側採樣（Gradient-based One-Side Sampling, GOSS）**：在每次迭代中，保留梯度較大的樣本（對訓練貢獻較大），並從梯度較小的樣本中隨機抽樣，藉此減少訓練資料量而不損失太多資訊量。
2. **獨佔特徵綑綁（Exclusive Feature Bundling, EFB）**：將互斥的稀疏特徵合併為單一特徵，降低有效特徵維度，加速分裂點的搜尋。

此外，LightGBM 採用**逐葉生長（Leaf-wise）** 策略，每次選擇增益最大的葉節點進行分裂，相較於 XGBoost 的逐層生長（Level-wise），在相同葉節點數下通常能達到更低的損失值，但也更容易過擬合，需搭配深度限制。

選用原因：

- 訓練效率高：GOSS 與 EFB 加速策略使其適合大規模資料
- 與 XGBoost 互補：同為梯度提升但建構策略不同，可比較兩種策略的效能差異
- 支援類別不平衡處理：透過 **is_unbalance** 或 **scale_pos_weight** 參數

實作參數：

- **n_estimators**: 100
- **max_depth**: -1（不限制，搭配 **num_leaves** 控制複雜度）
- **num_leaves**: 31
- **learning_rate**: 0.1
- **is_unbalance**: True（處理類別不平衡）

5.2.6 Support Vector Machine (SVM)

支援向量機（Support Vector Machine, SVM）由 Cortes & Vapnik (1995) 提出，基於統計學習理論，透過尋找最大間隔超平面進行分類。相較於樹模型依賴特徵的離散分割，SVM 在特徵空間中建立連續的決策邊界，對小樣本與高維資料具有較好的泛化能力。本研究採用 SVM 作為核方法（Kernel Method）的代表，與線性方法（LR、NB、LDA）、樹模型（DT、RF、XGBoost）及神經網路（MLP）形成四類方法的完整比較架構。

演算法原理：

SVM 尋找一個超平面 $w^T x + b = 0$，使得兩類樣本之間的間隔（margin）最大化。其目標為最小化 $\frac{1}{2} \| w \|^2$，使所有樣本 i 滿足：

$$y_i(w^T x_i + b) \geq 1 \tag{5-13}$$

對於非線性可分的情況，引入軟間隔（Soft Margin），允許部分樣本落在間隔內或錯誤分類，並以參數 C 控制分類錯誤的懲罰程度。

對於非線性問題，SVM 使用核函數（Kernel Function）將資料映射至高維空間。本研究採用徑向基函數（Radial Basis Function, RBF）核：

$$K(x_i, x_j) = \exp\left(-\gamma \| x_i - x_j \|^2\right) \tag{5-14}$$

RBF 核可隱式地將資料映射至無限維空間，使 SVM 能建立高度非線性的決策邊界。參數 γ 控制核函數的作用範圍：γ 越大，決策邊界越複雜。

選用原因：

- 理論基礎扎實：基於統計學習理論的結構風險最小化
- 適合中小型資料集：在樣本數有限時表現良好
- 支援 class_weight：可處理類別不平衡問題

實作參數：

- **kernel**: ‘rbf’（徑向基函數）
- **C**: 1.0
- **gamma**: ‘scale’
- **class_weight**: ‘balanced’

5.2.7 Multi-Layer Perceptron (MLP)

多層感知器（Multi-Layer Perceptron, MLP）是一種前饋神經網路（Feedforward Neural Network），透過多層神經元的非線性轉換學習複雜的特徵表示。MLP 的理論基礎可追溯至 Rumelhart et al. (1986) 提出的反向傳播（Backpropagation）演算法，該演算法使得多層神經網路的訓練成為可能。本研究使用 scikit-learn 的 MLPClassifier 實作。

演算法原理：

MLP 由輸入層、隱藏層與輸出層組成。本研究的網路架構為：

- 輸入層：26 個特徵（對應 26 個健檢指標與變化量）
- 隱藏層：2 層，每層 64 個神經元
- 輸出層：1 個神經元（二元分類）

前向傳播（Forward Propagation）：資料從輸入層逐層向前傳遞，每一層的計算包含線性轉換與非線性激活：

$$z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)} \tag{5-15}$$

$$a^{(l)} = \sigma(z^{(l)}) \tag{5-16}$$

其中 $W^{(l)}$ 與 $b^{(l)}$ 分別為第 l 層的權重矩陣與偏差向量，$a^{(l-1)}$ 為上一層的輸出（$a^{(0)} = X$ 為輸入特徵），σ 為激活函數。

本研究採用修正線性單元（Rectified Linear Unit, ReLU）作為隱藏層的激活函數：

$$\text{ReLU}(z) = \max(0, z) \tag{5-17}$$

相較於傳統的 sigmoid 激活函數，ReLU 具有兩項優勢：（1）計算效率高，僅需比較與取最大值操作；（2）緩解梯度消失問題（Vanishing Gradient Problem）——sigmoid 在輸入絕對值較大時梯度趨近於零，導致深層網路的參數難以更新，而 ReLU 在正值區域梯度恆為 1，使梯度能有效地向後傳播。

反向傳播與參數更新：訓練時，反向傳播演算法（Rumelhart et al., 1986）利用鏈式法則（Chain Rule）計算損失函數對每一層參數的梯度，再透過優化器更新權重。本研究採用 Adam 優化器，其結合了動量（Momentum）與自適應學習率（Adaptive Learning Rate），在神經網路訓練中被廣泛使用，相較於基本的隨機梯度下降（Stochastic Gradient Descent, SGD），Adam 對學習率的初始設定較不敏感，收斂速度更快。

選用原因：

- 非線性建模：透過多層非線性轉換，可學習複雜的特徵交互
- 彈性高：可調整網路深度與寬度以適應不同複雜度的任務
- 實作簡便：scikit-learn 提供統一的 API 介面

實作參數：

- **hidden_layer_sizes**: (64, 64)
- **activation**: ‘relu’

- **solver**: ‘adam’
- **max_iter**: 500

注意：MLPClassifier 不直接支援 class_weight 參數，本研究透過手動調整樣本權重（sample_weight）處理類別不平衡問題。

5.3 符號回歸

5.3.1 PySR

符號回歸（Symbolic Regression）透過遺傳規劃（Genetic Programming, GP）演化出可解釋的數學公式（Cranmer, 2023）。相較於前述的黑盒模型，符號回歸產出的公式可直接理解其醫學意義，兼顧預測能力與可解釋性。

演算法原理：

遺傳規劃的核心流程為：

1. **初始化**：隨機生成一群數學公式（個體），每個公式為一棵運算樹
2. **適應度評估**：計算每個公式的預測誤差（如均方誤差）
3. **選擇**：保留表現較好的公式
4. **演化操作**：透過交叉（Crossover，交換子樹）與突變（Mutation，隨機修改節點）產生新公式
5. **迭代**：重複步驟 2-4 直到收斂

PySR 在搜尋過程中同時考慮公式的精度與複雜度，產出一系列 Pareto 前沿上的候選公式，供研究者依據領域知識選擇最適當的公式。

選用原因：

- 完全透明：產出的公式可直接理解
- 領域知識驗證：可檢驗公式是否符合醫學邏輯
- 輕量部署：簡單公式不需複雜運算資源

使用套件：本研究初期使用 Python 原生的 gplearn 套件進行符號回歸實驗，但因其不支援 class_weight 且搜尋效率有限，後改用基於 Julia 的 PySR 套件（Cranmer,

2023）。PySR 支援 sample_weight 且搜尋效能更佳，為本研究最終採用之符號回歸工具。

5.4 類別不平衡處理策略

由於三高疾病的患病率較低（高血壓 16.68%、高血糖 5.53%、高血脂 5.96%），本研究採用 class_weight='balanced' 作為主要的類別不平衡處理策略（原理與公式詳見第二章）。各模型的具體設定如下：

- **LR、DT、RF、SVM**：設定 **class_weight='balanced'**，由 scikit-learn 自動依類別頻率倒數計算權重
- **XGBoost**：設定 **scale_pos_weight** 為負正類比例，效果等同 balanced
- **LightGBM**：設定 **is_unbalance=True**，自動調整類別權重
- **MLP**：透過手動調整 **sample_weight** 實現加權訓練
- **NB、LDA、KNN**：不支援 class_weight，以原始資料分佈訓練
- **PySR**：透過 **sample_weight** 參數加權

第六章 實驗結果

本章呈現各項實驗的結果，包括模型性能比較、特徵重要性分析、消融實驗以及符號回歸實驗。所有實驗皆採用滑動窗口資料集（13,514 筆紀錄）與 StratifiedGroupKFold 5-fold 交叉驗證，確保同一參與者的紀錄不會同時出現在訓練集與測試集中。

6.1 模型性能比較

實驗目的：回答 Q2（模型選擇與比較）——在三高疾病預測任務中，哪些機器學習模型表現最佳？傳統統計方法與機器學習方法的性能差異為何？

6.1.1 整體結果

表 6-1 呈現十種模型在三項預測任務上的 AUC 表現。模型涵蓋傳統統計方法（LR、NB、LDA）、基於實例方法（KNN）、樹模型（DT、RF、XGB、LGBM）、核方法（SVM）及神經網路（MLP）。整體而言，Logistic Regression 在高血糖與高血脂預測中皆達到最高 AUC，尤其高血糖預測達 0.938；高血壓預測則以 Random Forest 表現最佳（AUC 0.743）。

表 6-1 各模型 AUC 比較（5-Fold CV）

模型	類型	高血壓	高血糖	高血脂
LR	傳統統計	0.721 ± 0.017	**0.938 ± 0.010**	**0.867 ± 0.012**
NB	傳統統計	0.709 ± 0.022	0.917 ± 0.010	0.847 ± 0.015
LDA	傳統統計	0.720 ± 0.017	0.936 ± 0.011	0.867 ± 0.012
KNN	基於實例	0.630 ± 0.018	0.782 ± 0.020	0.673 ± 0.013
DT	樹模型	0.658 ± 0.012	0.835 ± 0.014	0.744 ± 0.037
RF	樹模型	**0.743 ± 0.013**	0.932 ± 0.008	0.859 ± 0.014
XGB	樹模型	0.738 ± 0.012	0.930 ± 0.014	0.857 ± 0.016
LGBM	樹模型	0.730 ± 0.011	0.926 ± 0.015	0.852 ± 0.011
SVM	核方法	0.726 ± 0.011	0.919 ± 0.012	0.845 ± 0.012
MLP	神經網路	0.703 ± 0.033	0.919 ± 0.021	0.742 ± 0.136

註：粗體表示該疾病最佳結果；所有數值為 mean ± std

圖 6-1 呈現八種模型在三項疾病預測任務上的 ROC 曲線。由圖可知，高血糖預測的 ROC 曲線整體最靠近左上角，高血脂次之，高血壓最低——與表 6-1 的 AUC 數值趨勢一致。

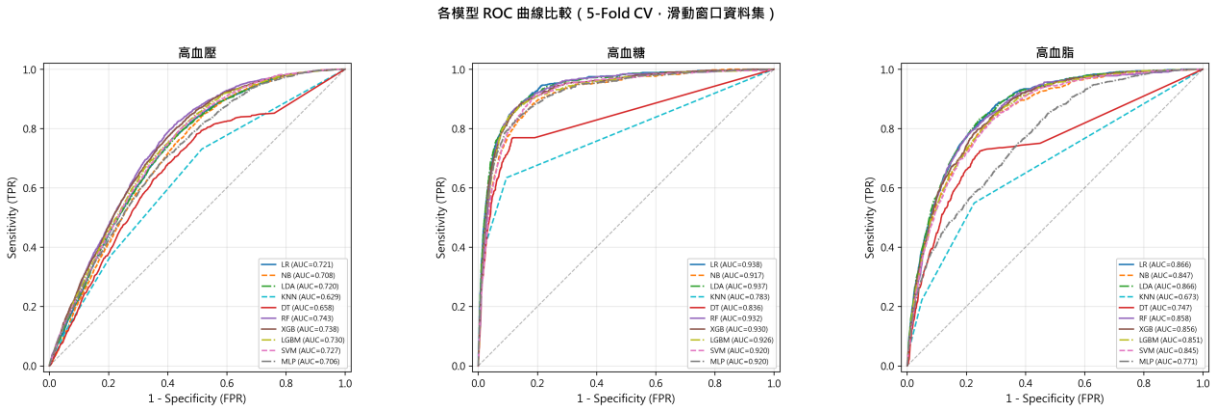

圖 6-1 各模型 ROC 曲線比較（5-Fold CV）

6.1.2 高血壓預測結果

高血壓預測（陽性率 16.68%）的 AUC 介於 0.630 至 0.743 之間。Random Forest 達到最高的 AUC（0.743），其次為 XGBoost（0.738）與 SVM（0.726）。傳統統計方法中，LR（0.721）與 LDA（0.720）表現接近，NB 略低（0.709）。KNN 表現最差（0.630），而 MLP 則呈現較大的變異（標準差 0.033）。

表 6-2 呈現高血壓預測的完整評估指標。值得注意的是，RF 雖然 AUC 最高，但其 Sensitivity 僅 0.286，顯示模型傾向保守預測。LDA 與 MLP 皆呈現極端的保守預測行為（Sensitivity 分別僅 0.037 與 0.017），幾乎將所有樣本判為非患病。NB 的 Sensitivity（0.357）雖高於 LDA，但仍明顯低於 LR（0.697）。相較之下，LR 與 SVM 在 Sensitivity 與 Specificity 之間取得較佳的平衡。

表 6-2 高血壓預測詳細結果

模型	AUC	Sensitivity	Specificity	F1-Score
LR	0.721	0.697	0.638	0.434
NB	0.709	0.357	0.832	0.347
LDA	0.720	0.037	0.988	0.068
KNN	0.630	0.116	0.946	0.172
DT	0.658	0.646	0.629	0.404
RF	**0.743**	0.286	0.890	0.328
XGB	0.738	0.678	0.676	0.447
LGBM	0.730	0.601	0.717	0.432
SVM	0.726	0.704	0.635	0.436
MLP	0.703	0.017	0.996	0.032

6.1.3 高血糖預測結果

高血糖預測（陽性率 5.53%）展現最佳的預測效能，除 KNN（0.782）外，其餘模型的 AUC 皆高於 0.83。Logistic Regression 達到最高的 AUC（0.938），其次為 LDA（0.936）、RF（0.932）、XGBoost（0.930）與 LightGBM（0.926）。

表 6-3 高血糖預測詳細結果

模型	AUC	Sensitivity	Specificity	F1-Score
LR	**0.938**	0.858	0.882	0.461
NB	0.917	0.601	0.953	0.511
LDA	0.936	0.484	0.980	0.536
KNN	0.782	0.223	0.992	0.331
DT	0.835	0.763	0.887	0.431
RF	0.932	0.551	0.965	0.525
XGB	0.930	0.745	0.933	0.532
LGBM	0.926	0.646	0.954	0.542
SVM	0.919	0.721	0.922	0.488
MLP	0.919	0.304	0.988	0.399

表 6-3 進一步呈現高血糖預測的詳細指標。LR 在 AUC 最高的同時，Sensitivity 亦達 0.858，顯示其能有效識別大多數患病者。相較之下，LDA 雖 AUC 接近（0.936），但 Sensitivity 僅 0.484，約半數患者被漏判。KNN 與 MLP 的 Sensitivity 分別僅 0.223 與 0.304，臨床實用性有限。在 F1-Score 方面，LGBM（0.542）與 LDA（0.536）表現

最佳，反映其在 Precision 與 Recall 間取得較好的平衡。整體而言，高血糖因陽性率最低（5.53%），各模型的 F1-Score 普遍偏低（0.331–0.542），顯示在極度不平衡的情境下，單純以 AUC 評估可能高估模型的臨床效用。

6.1.4 高血脂預測結果

高血脂預測（陽性率 5.96%）的 AUC 介於 0.673 至 0.867 之間。LR 與 LDA 並列最高 AUC（0.867），RF（0.859）、XGBoost（0.857）與 LightGBM（0.852）緊隨其後。KNN 表現最差（0.673），而 MLP 呈現最大的不穩定性（標準差 0.136）。

表 6-4 高血脂預測詳細結果

模型	AUC	Sensitivity	Specificity	F1-Score
LR	**0.867**	0.799	0.775	0.362
NB	0.847	0.416	0.941	0.396
LDA	**0.867**	0.118	0.991	0.193
KNN	0.673	0.061	0.992	0.105
DT	0.744	0.673	0.785	0.323
RF	0.859	0.391	0.942	0.378
XGB	0.857	0.676	0.844	0.388
LGBM	0.852	0.561	0.891	0.397
SVM	0.845	0.695	0.821	0.368
MLP	0.742	0.066	0.995	0.108

高血脂預測呈現與高血壓類似的 Sensitivity 極化現象。LR 的 Sensitivity 最高（0.799），其次為 SVM（0.695）與 XGBoost（0.676），而 LDA（0.118）、KNN（0.061）與 MLP(0.066)幾乎喪失識別患病者的能力。值得注意的是，LDA 雖與 LR 並列最高 AUC（0.867），但 Sensitivity 相差近 7 倍（0.118 vs 0.799），凸顯 AUC 作為單一指標的局限性。MLP 的標準差高達 0.136，遠超其他模型（均 < 0.04），顯示其在此預測任務中的訓練不穩定。

6.2 特徵重要性分析

實驗目的：回答 Q4（特徵重要性分析）——哪些生物標記及其變化量對三高疾病預測最為重要？Δ 特徵在重要特徵中的佔比為何？

6.2.1 SHAP 特徵重要性

本研究使用 SHAP（SHapley Additive exPlanations）分析 XGBoost 模型的特徵重要性。

表 6-5 各疾病 Top 10 重要特徵（SHAP）

排名	高血壓	高血糖	高血脂
1	SBP_Y-2	FBG_Y-1	TC_Y-2
2	SBP_Y-1	FBG_Y-2	TC_Y-1
3	Age	ΔTC	ΔeGFR
4	ΔDBP	BMI_Y-1	Age
5	DBP_Y-1	BMI_Y-2	ΔTC
6	DBP_Y-2	ΔUA	DBP_Y-1
7	ΔeGFR	ΔeGFR	ΔFBG
8	ΔSBP	ΔFBG	ΔUA
9	eGFR_Y-2	eGFR_Y-2	BMI_Y-1
10	BMI_Y-1	UA_Y-1	ΔBMI

由表 6-5 可觀察到，各疾病的最重要特徵與其直接相關的生物標記高度吻合：高血壓以收縮壓（SBP）與舒張壓（DBP）為主、高血糖以空腹血糖（FBG）為主、高血脂以總膽固醇（TC）為主。此外，同一指標的兩個時間點（Y-2 與 Y-1）通常同時出現在 Top 10 中，顯示模型同時參考歷史趨勢與近期數值。年齡（Age）在高血壓與高血脂中皆進入前四名，反映年齡作為慢性病共通風險因子的角色。ΔeGFR 則同時出現在三項疾病的 Top 10 中，暗示腎功能變化可能是三高疾病的共通預測指標。

圖 6-2 以水平長條圖呈現三項疾病的 Top 10 重要特徵比較，可直觀觀察各特徵對不同疾病的相對重要程度。

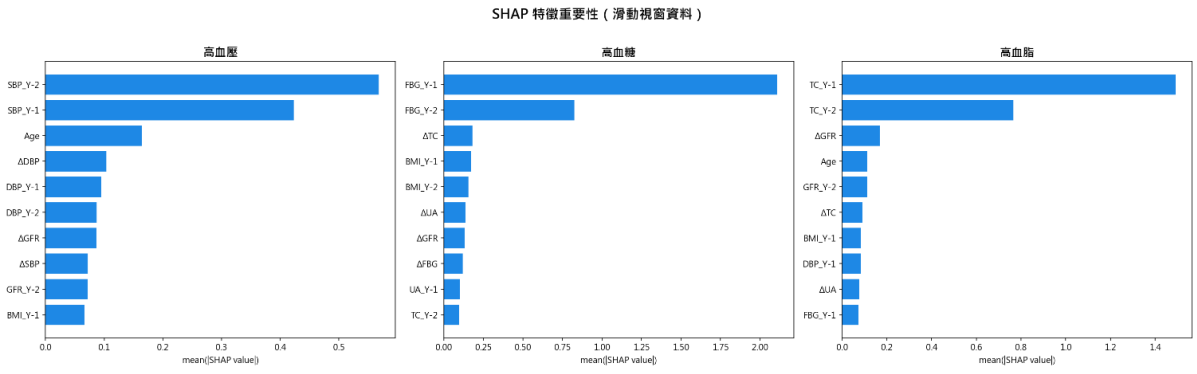

圖 6-2 三項疾病 SHAP 特徵重要性比較（XGBoost, Top 10）

6.2.2 Δ 特徵在 Top 10 中的佔比

在 Top 10 重要特徵中，Δ 特徵的佔比如表 6-6 所示：

表 6-6 各疾病 Top 10 中 Δ 特徵數量

疾病	Δ 特徵數量	佔比
高血壓	3 (ΔDBP, ΔeGFR, ΔSBP)	30%
高血糖	4 (ΔTC, ΔUA, ΔeGFR, ΔFBG)	40%
高血脂	5 (ΔeGFR, ΔTC, ΔFBG, ΔUA, ΔBMI)	50%

Δ 特徵在 Top 10 中的佔比從高血壓的 30% 逐步上升至高血脂的 50%，顯示變化量特徵對高血脂預測的相對重要性最高。此趨勢與 6.3 節消融實驗中 Δ 特徵對高血脂提升最大（+2.1%）的結果呈現一致方向。值得注意的是，ΔeGFR 為唯一同時出現在三項疾病 Top 10 的 Δ 特徵，呼應腎功能變化作為代謝異常早期指標的臨床觀點。

圖 6-3 以高血壓為例，呈現 SHAP beeswarm 圖。每個點代表一筆樣本，橫軸為 SHAP 值（對預測的影響方向與大小），顏色代表特徵值的高低。

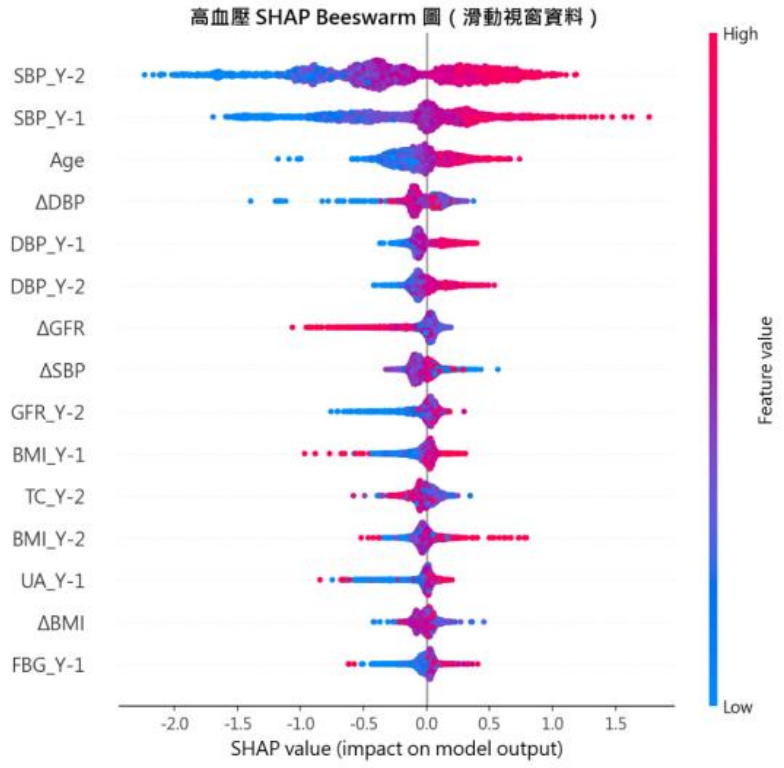

圖 6-3 高血壓預測 SHAP Beeswarm 圖（XGBoost）

6.3 Δ 特徵消融實驗

實驗目的：回答 Q1（變化量特徵的預測價值）——額外納入 Δ 特徵是否能顯著提升三高疾病的預測性能？

6.3.1 實驗設計

為驗證 Δ 特徵的預測價值，本研究設計兩組比較框架：

1. **比較框架 1**：Full (Y-2 + Y-1 + Δ) vs No-Delta (Y-2 + Y-1)
    - 目的：在完整特徵集中移除 Δ 特徵的影響
2. **比較框架 2**：Y-1 + Δ vs Y-1 Only
    - 目的：在僅有單一時間點資料時，評估 Δ 特徵的增量價值

6.3.2 比較框架 1 結果

移除 Δ 特徵後，三種疾病的 AUC 均無變化（表 6-7），顯示在已包含 Y-2 與 Y-1 兩個時間點靜態特徵的情況下，Δ 特徵未能提供額外的預測資訊。此結果在統計上合理：由於 Δ = Y-1 − Y-2，LR 等線性模型可直接從 Y-2 與 Y-1 的係數差隱含地計算出等效的變化量效果，因此 Δ 特徵在 Full 特徵集中屬於冗餘資訊。

表 6-7 Full vs No-Delta 比較（LR 模型）

疾病	Full (26 特徵)	No-Delta (18 特徵)	差異
高血壓	0.721	0.721	0.0%
高血糖	0.938	0.938	0.0%
高血脂	0.867	0.867	0.0%

6.3.3 比較框架 2 結果

當模型僅有 Y-1 資料時，加入 Δ 特徵可帶來 1.5%–2.3% 的 AUC 提升（表 6-8）。三項疾病均呈現一致的改善趨勢，其中高血壓的提升幅度最大（+2.3%），顯示在缺乏 Y-2 靜態特徵的情況下，Δ 特徵能有效補充時序變化資訊。此結果對僅有單次健檢紀錄的臨床場景具有實務意義：透過計算兩次健檢間的變化量，即可在不增加健檢項目的前提下提升預測效能。

表 6-8 Y-1 + Δ vs Y-1 Only 比較（LR 模型）

疾病	Y-1 + Δ (18 特徵)	Y-1 Only (10 特徵)	提升
高血壓	0.721	0.698	**+2.3%**
高血糖	0.938	0.923	**+1.5%**
高血脂	0.867	0.846	**+2.1%**

圖 6-4 以長條圖呈現 LR 與 XGBoost 兩種模型在 Full vs No-Delta 條件下的 AUC 比較。兩種模型在三項疾病中均呈現相似的模式：移除 Δ 特徵對 AUC 的影響極小，進一步支持比較框架 1 的結論。

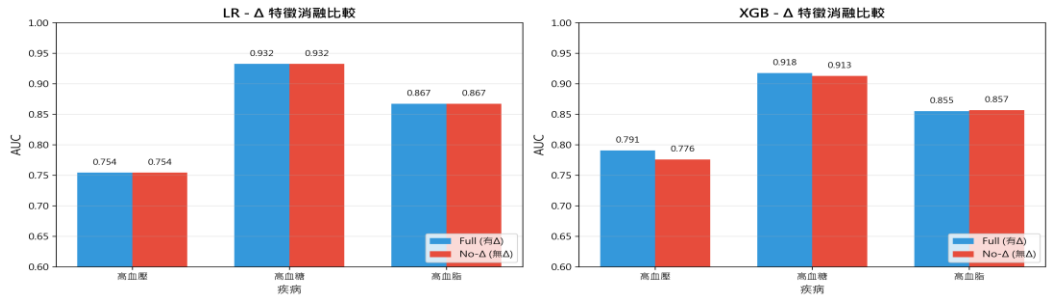

圖 6-4 Δ 特徵消融實驗結果（Full vs No-Delta）

6.4 特徵選擇消融實驗

實驗目的：回答 Q5（模型精簡的可行性）——使用少量關鍵特徵是否能維持接近完整模型的預測性能？

6.4.1 實驗設計

為評估精簡特徵集的可行性，本研究依據 SHAP 特徵重要性排序，測試使用 Top 5、Top 10、Top 15、Top 20 與全部 26 個特徵的預測效能。

6.4.2 實驗結果

由表 6-9 可觀察到，僅使用 Top 5 特徵即可達到與全部 26 個特徵幾乎相同的 AUC。以 LR 為例，高血壓的 Top 5 AUC（0.752）與全特徵 AUC（0.754）僅差 0.002，高血糖與高血脂更分別達到 0.933 與 0.868，甚至略高於全特徵結果。XGBoost 亦呈現類似趨勢，Top 5 至 Top 26 的 AUC 差異均在 0.005 以內。此結果顯示，少量關鍵特徵即可維持模型的預測效能，為臨床場景中的精簡篩檢方案提供了實證基礎。

表 6-9 特徵選擇消融實驗結果

特徵數	高血壓 (LR)	高血壓 (XGB)	高血糖 (LR)	高血糖 (XGB)	高血脂 (LR)	高血脂 (XGB)
5	0.752	0.785	0.933	0.913	0.868	0.857
10	0.754	0.784	0.933	0.911	0.868	0.854
15	0.755	0.784	0.932	0.910	0.867	0.850
20	0.755	0.786	0.932	0.917	0.867	0.855
26	0.754	0.787	0.932	0.915	0.867	0.859

圖 6-5 以折線圖呈現不同特徵數量下 LR 與 XGBoost 的 AUC 變化趨勢，虛線標示全特徵 AUC 的 95% 門檻。由圖中三項任務在 Top 5 即已達到或接近該門檻。

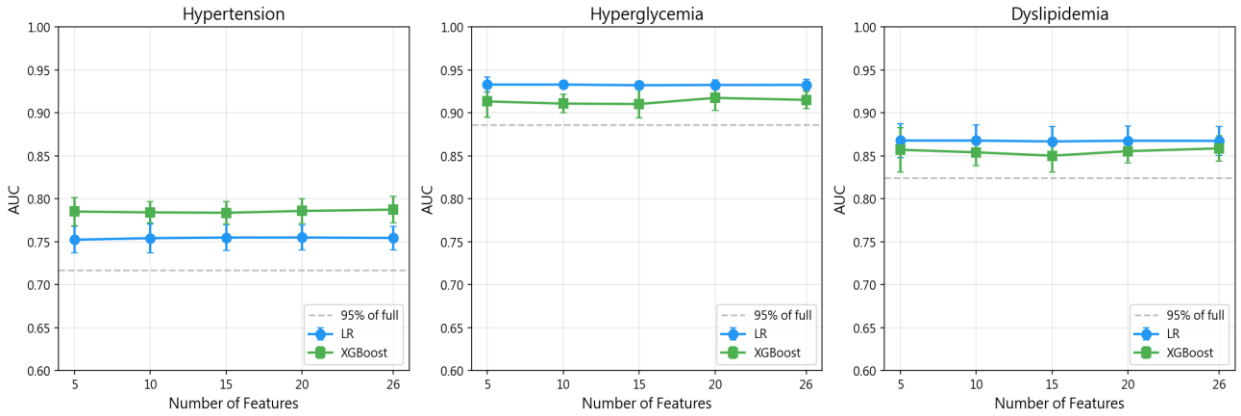

圖 4-5 特徵數量與 AUC 關係（LR vs XGBoost）

6.5 類別不平衡處理比較

實驗目的：驗證類別不平衡處理策略的穩健性——比較 class_weight、SMOTE、ADASYN 等方法對模型效能的影響，選定最適合實務應用的方案。

6.5.1 實驗設計

由於三高疾病的陽性率介於 5.5% 至 16.7%，類別不平衡可能影響模型效能。本研究分兩階段進行比較：

1. **方法比較**：比較五種類別不平衡處理方法（Baseline、class_weight、SMOTE、ADASYN、RandomUnderSampler）。
2. **權重設定比較**：針對 class_weight 方法，進一步比較不同權重設定（None、balanced、1:3、1:5、1:10）對 Sensitivity-Specificity 權衡的影響。

6.5.2 方法比較結果

表 6-10 呈現五種類別不平衡處理方法的 AUC 與 Sensitivity 比較。

表 6-10 類別不平衡處理方法比較（LR 模型）

疾病	指標	Baseline	class_weight	SMOTE	ADASYN	Under-Sampling
高血壓	AUC	0.719	0.720	0.719	0.719	0.719
	Sens	0.041	0.698	0.698	0.696	0.699
高血糖	AUC	0.937	0.937	0.937	0.936	0.937
	Sens	0.335	0.861	0.852	0.877	0.864
高血脂	AUC	0.863	0.864	0.863	0.863	0.862
	Sens	0.135	0.791	0.785	0.794	0.790

五種方法的主要發現：

1. **AUC 幾乎無差異**：所有方法的 AUC 差異小於 0.2%，顯示不同的不平衡處理策略對模型的排序能力無顯著影響。
2. **Sensitivity 大幅改善**：相較於 Baseline（未處理），四種處理方法皆可將 Sensitivity 從極低值（0.04-0.34）提升至 0.70-0.88，效果近乎等同。
3. **方法間差異極小**：class_weight、SMOTE、ADASYN 與 RandomUnderSampler 四種方法的 Sensitivity 差異 < 2%，而 class_weight 不需生成合成樣本、不改變資料分佈，為最簡便的實務選擇。

6.5.3 權重設定比較

在確認 class_weight 為適當策略後，進一步比較不同權重設定的影響。

表 6-11 不同 class_weight 設定比較（LR 模型）

疾病	指標	None	balanced	1:3	1:5	1:10
高血壓	AUC	0.754	0.754	0.754	0.754	0.754
	Sens	0.053	0.744	0.513	0.744	0.903
高血糖	AUC	0.933	0.932	0.934	0.934	0.933
	Sens	0.367	0.854	0.579	0.690	0.764
高血脂	AUC	0.867	0.867	0.868	0.868	0.867
	Sens	0.094	0.806	0.349	0.515	0.693

由表 6-11 與圖 6-6 可見，不同 class_weight 設定呈現清晰的 Sensitivity-Specificity 權衡：從 None 到 1:10，Sensitivity 逐步上升而 Specificity 下降，AUC 則維持不變（差異 < 0.2%）。balanced 設定在三種疾病中皆取得 Sensitivity 0.74-0.85 的水準，同時維持合理的 Specificity，適合疾病篩檢應用場景。

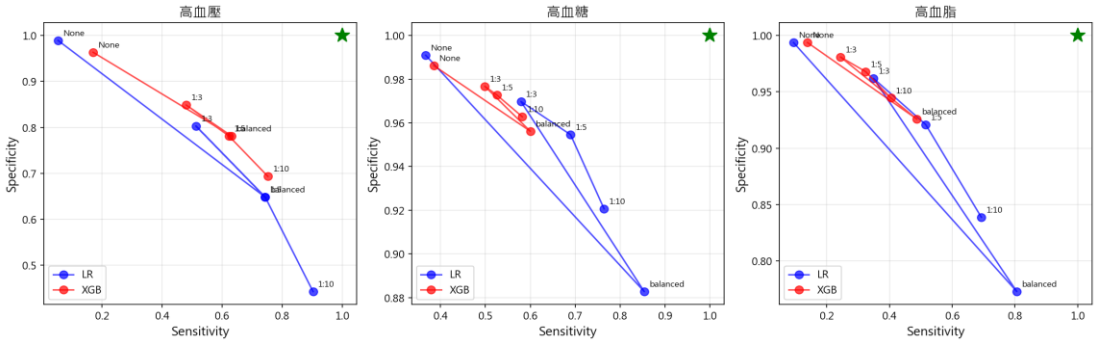

圖 6-6 不同 class_weight 設定下的 Sensitivity-Specificity 權衡（LR 模型）

6.6 資料篩選策略比較

實驗目的：驗證資料處理策略的穩健性——排除基線已確診個案是否影響預測效能？確保主要實驗結果不受資料篩選方式左右。

6.6.1 實驗設計

資料集中部分滑動窗口樣本的起點（Y-2）已有目標疾病，這些「非新發」個案是否應該排除？為回答此問題，本研究設計三組比較策略：

- **策略 A（目前做法）**：包含所有滑動窗口樣本（13,514 筆）。
- **策略 B**：排除首次健檢（Times=1）已確診的個案。
- **策略 C**：排除每個滑動窗口中，Y-2 已有目標疾病的樣本。

診斷閾值依據 Luo et al. (2024)：高血壓 SBP ≥ 140 或 DBP ≥ 90 mmHg、高血糖 FBG ≥ 7.0 mmol/L、高血脂 TC ≥ 6.22 mmol/L。由於原始資料集在首次健檢（Times=1）時疾病率幾乎為 0%（Luo et al. 2024 已隱含基線健康篩選），策略 B 與策略 A 實質上無差異，因此本實驗聚焦於比較策略 A 與策略 C。

6.6.2 排除統計

表 6-12 呈現各疾病在策略 C 下的排除統計（原始樣本數皆為 13,514 筆）。高血壓因持續患病者較多，排除比例最高（10.4%）；高血糖與高血脂的排除比例較小。排除後陽性率均略降，反映移除了部分「持續患病」的正類樣本。

表 6-12 策略 C 排除統計

疾病	Y-2 已確診	排除後樣本數	原始陽性率	排除後陽性率
高血壓	1,402 (10.4%)	12,112	19.3%	17.9%
高血糖	367 (2.7%)	13,147	5.9%	4.6%
高血脂	548 (4.1%)	12,966	7.9%	6.8%

6.6.3 實驗結果

表 6-13 排除策略 AUC 比較（5-Fold CV）

疾病	模型	A（全部樣本）	C（排除已確診）	差異
高血壓	LR	0.712 ± 0.015	0.710 ± 0.011	-0.3%
高血壓	RF	0.735 ± 0.012	0.748 ± 0.014	+1.3%
高血糖	LR	0.922 ± 0.017	0.910 ± 0.018	-1.1%
高血糖	RF	0.924 ± 0.011	0.917 ± 0.014	-0.8%
高血脂	LR	0.858 ± 0.012	0.854 ± 0.011	-0.4%
高血脂	RF	0.851 ± 0.017	0.844 ± 0.012	-0.6%

由表 6-13 可見，兩種策略的 AUC 差異均在 1.3% 以內，且無一致性方向（高血壓 RF 略升 +1.3%，其餘皆略降）。此結果驗證了本研究採用策略 A（包含所有樣本）的合理性：排除基線已確診個案並未帶來系統性的效能改善，而保留全部樣本可維持較大的訓練資料量與統計效力。

6.7 健檢次數與預測效能

實驗目的：回答 Q6（健檢次數對預測性能的影響）——累積更多次健檢紀錄是否能提升預測準確度？

6.7.1 實驗設計

為探討縱向健檢資料的累積效益，本研究設計健檢次數比較實驗。固定以預測年（Y）作為預測目標，逐步增加輸入的歷史健檢次數：1 次（僅 Y-1）、2 次（Y-2、Y-1）、

3 次（Y-3～Y-1）、4 次（Y-4～Y-1），比較不同健檢次數下 LR 模型的 AUC 表現。為確保公平比較，四組實驗使用相同的 2,526 名參與者（即具有完整 5 次健檢紀錄者）。

6.7.2 實驗結果

圖 4-7 呈現不同健檢次數下三項疾病的 AUC 比較。

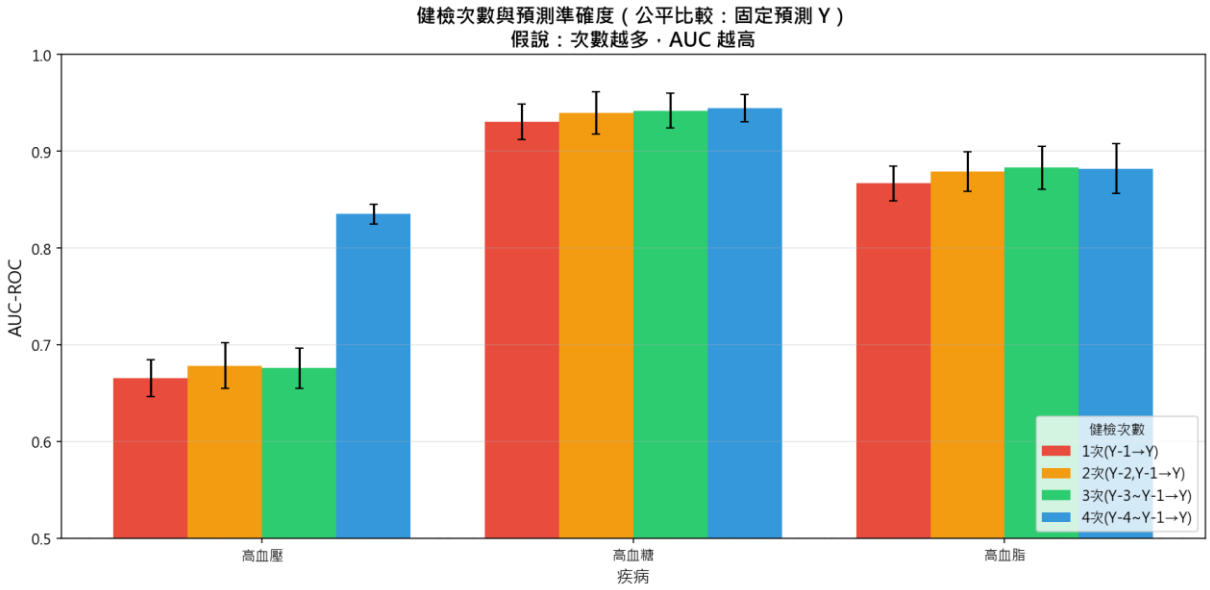

圖 6-7 健檢次數與預測準確度（LR 模型，固定預測 Y）

由圖 6-7 可觀察到，高血糖與高血脂的 AUC 隨健檢次數增加呈現緩步上升趨勢，但增幅有限（約 1–2%）。高血壓的表現較為特殊：1 至 3 次健檢的 AUC 維持在約 0.67–0.68 的水準，但在 4 次健檢時出現明顯跳升至約 0.84，顯示更長期的血壓變化趨勢對高血壓預測具有較大的增量價值。整體而言，2 次健檢（即本研究的主要實驗設定）已能提供穩定的預測基礎，但累積更多健檢紀錄仍有潛在的效能提升空間，尤其是高血壓預測。

6.8 多任務學習與單任務學習比較

實驗目的：回答 Q3（多任務學習的效果）——同時預測三高疾病是否優於分別預測單一疾病？

6.8.1 實驗設計

本研究的三項預測目標（高血壓、高血糖、高血脂）皆來自相同的健檢資料與特徵集，適合以多任務學習（Multi-Task Learning, MTL）架構同時預測。為評估 MTL 是否優於獨立訓練三個模型的單任務學習（Single-Task Learning, STL），本研究設計以下比較實驗：

- **MTL 架構**：共享底層神經網路（64 → 32 節點），頂部設置三個獨立輸出頭，分別對應三項疾病。損失函數為三項任務的加權平均。
- **STL 架構**：針對每項疾病各自訓練一個獨立的 MLP 模型，網路結構與 MTL 的單任務路徑相同（64 → 32 → 1）。

兩組實驗使用相同的滑動窗口資料集與 StratifiedGroupKFold 5-fold 交叉驗證。

6.8.2 實驗結果

表 6-14 MTL vs STL 比較（MLP 模型）

疾病	MTL	STL	差異
高血壓	0.734 ± 0.020	0.742 ± 0.020	-0.8%
高血糖	0.932 ± 0.007	0.933 ± 0.007	-0.1%
高血脂	0.868 ± 0.009	0.869 ± 0.010	-0.1%
平均	**0.845**	**0.848**	**-0.3%**

MTL 與 STL 的 AUC 差異極小（平均僅 -0.3%），且三項疾病的差異方向一致（MTL 皆略低於 STL）。此結果顯示，三高疾病雖共享相同的健檢特徵，但透過共享神經網路底層並未產生正向的遷移學習效果。可能的原因是三項疾病的主要預測特徵差異較大（如高血壓以 SBP/DBP 為主、高血糖以 FBG 為主），共享表徵反而略微稀釋了各任務的專屬資訊。基於此結果，本研究的主要實驗採用 STL 架構，分別訓練三項疾病的獨立模型。

6.9 符號回歸實驗

實驗目的：探索可解釋的數學公式——能否從資料中發現簡潔的預測公式，兼顧預測效能與臨床可解釋性？

6.9.1 實驗設計

符號回歸使用 PySR 套件，嘗試從資料中自動發現可解釋的數學公式。實驗設定：最大複雜度 20、迭代次數 100、運算子包含 +、-、*、/、exp。

6.9.2 發現的公式

表 6-15 符號回歸發現的公式

疾病	公式	AUC	訓練時間
高血壓	$0.130 \times \exp(SBP_{Y-2})$	0.745	20.7 分鐘
高血糖	$0.114 \times FBG_{Y-1}$	0.943	20.6 分鐘
高血脂	$0.043 \times \exp(TC_{Y-2})$	0.801	20.8 分鐘

PySR 發現的公式呈現高度簡潔性：三項疾病的預測公式均僅依賴單一特徵，且該特徵恰好是 SHAP 分析中排名第一的重要特徵（高血壓：SBP_Y-2、高血糖：FBG_Y-1、高血脂：TC_Y-2）。高血糖的公式最為簡單——直接以 FBG_Y-1 的線性函數預測，即可達到 0.943 的 AUC，甚至略高於 XGBoost（0.930）。高血壓與高血脂的公式則使用指數函數，反映這兩項疾病與對應生物標記之間的非線性關係。從臨床可解釋性的角度，這些公式直觀地呈現了「血壓高→高血壓風險高」、「血糖高→高血糖風險高」的醫學常識，驗證了模型學習到的特徵與臨床知識一致。然而，高血脂公式的 AUC（0.801）較完整模型（0.867）低 7.6%，顯示單一特徵在此任務中的預測力有限。

第七章 討論

本章針對第六章的實驗結果進行深入討論與詮釋，分析各實驗發現的意義，並探討其對臨床應用的啟示。

7.1 模型選擇與效能分析

7.1.1 Logistic Regression 的穩定優勢

綜合十種模型在三項預測任務的結果，Logistic Regression 表現最為穩定：在三項任務中皆達到最高或接近最高的 AUC，且 Sensitivity/Specificity 平衡良好，適合作為臨床應用的基準模型。此結果與 Sun et al. (2017) 系統性回顧中 LR 長期作為疾病預測主流方法的觀察一致。

7.1.2 傳統統計方法的表現差異

LDA 在三項任務中的 AUC 與 LR 幾乎一致（高血壓 0.720 vs 0.721、高血糖 0.936 vs 0.938、高血脂 0.867 vs 0.867），然而其 Sensitivity 極低（高血壓 0.037、高血脂 0.118），顯示 LDA 在類別不平衡資料中傾向將樣本歸類為多數類。此結果與 LDA 的決策邊界特性一致——當類別先驗機率差異懸殊時，判別邊界偏向少數類方向。

NB 在三項任務中的 AUC 均低於 LR 約 1~2%，其特徵獨立性假設在生理指標間存在相關性（如 SBP 與 DBP、BMI 與血脂）的情境下，限制了模型的判別能力。

7.1.3 集成模型與神經網路

Random Forest 與 XGBoost 表現相近，兩者皆為強健的集成方法，但傾向保守預測（高 Specificity、低 Sensitivity）。MLP 穩定性不足，在高血壓與高血脂任務中呈現極端的預測行為（幾乎全部預測為非患病），標準差亦較大。LightGBM 表現接近 XGBoost，驗證了梯度提升框架在此任務的一致性。KNN 在所有任務中表現最差（AUC 0.630-0.782），因其基於距離的決策方式在高維特徵空間中效率偏低，且不支援 class_weight 調整，導致在不平衡資料上 Sensitivity 極低。Decision Tree 預測效能亦偏低，單一決策樹容易過擬合，不建議作為最終預測模型。

7.1.4 疾病間預測難度差異

三項任務中，高血糖的 AUC 普遍最佳（除 KNN 外，所有模型 AUC > 0.83），可能因 FBG 本身即為診斷指標的直接前驅，具有極高的預測價值。高血壓的預測效能相對較低（AUC 0.630-0.743），反映血壓的波動性較大，且受更多環境因素影響。

整體而言，簡單的線性模型在結構化健檢資料上已具備與複雜模型相當的排序能力，此發現對臨床部署具有重要意義——無需複雜的模型即可達到實用的預測效能。

7.2 特徵工程的價值

7.2.1 疾病特異性預測因子

SHAP 特徵重要性分析清楚呈現了疾病特異性：高血壓預測以血壓相關特徵（SBP、DBP）為主導；高血糖預測以空腹血糖（FBG）及代謝相關指標為主；高血脂預測則以總膽固醇（TC）與腎功能變化（ΔeGFR）最為重要。這些發現符合臨床直覺，也為個人化風險評估與健康管理建議提供了明確的指引。

7.2.2 Δ 特徵的雙重角色

Δ 特徵消融實驗揭示了一個重要的雙重特性：

1. **資訊冗餘**：當同時有 Y-2 與 Y-1 時，Δ 特徵為冗餘資訊（Δ = Y-1 − Y-2），模型可自行學習變化量，故移除後 AUC 不變。
2. **資料受限情境的價值**：當僅有 Y-1 資料時，Δ 特徵可提供約 2% 的 AUC 提升（高血壓 +2.3%、高血脂 +2.1%、高血糖 +1.5%），展現其實用價值。Δ 特徵用 8 個變數編碼了兩個時間點的關係資訊，達到特徵壓縮的效果。

此發現與 Yang et al. (2025) 中 δ-FPG 特徵的高重要性相呼應，為縱向特徵工程在慢性疾病預測的應用提供了跨疾病的實證支持。

7.2.3 特徵精簡的可行性

特徵選擇消融實驗顯示，僅使用 SHAP 排序前 5 個重要特徵，LR 模型在三項任務中的 AUC 降幅皆小於 0.5%。XGBoost 對特徵數量略為敏感，但差異同樣有限。此結果證實精簡模型在維持預測性能的同時可大幅降低資料收集成本，對基層醫療單位的實務部署具有重要意義。

7.3 類別不平衡與資料策略

7.3.1 不平衡處理方法的等價性

五種類別不平衡處理方法（Baseline、class_weight、SMOTE、ADASYN、RandomUnderSampler）的 AUC 差異小於 0.2%，顯示不同的不平衡處理策略對模型的排序能力無顯著影響。然而，四種處理方法皆可將 Sensitivity 從極低值（0.04-0.34）提升至 0.70-0.88，效果近乎等同。

class_weight（balanced 設定）為最佳實務選擇：效果等同於 SMOTE 等過採樣方法，但實作更簡單、無需生成合成樣本、不改變資料分佈。balanced 設定在維持 AUC 的同時，Sensitivity 可提升至 0.74-0.85，適合疾病篩檢應用。

7.3.2 資料篩選策略的穩健性

排除窗口起點已確診樣本後，AUC 變化極小（|差異|≤1.3%）。六組比較中，五組 AUC 略微下降，僅 RF 在高血壓任務中略微上升（+1.3%）。AUC 下降的合理解釋為：移除窗口起點已確診個案等於移除「容易預測」的正類樣本（持續患病者），使預測任務聚焦於更困難的「新發」個案。

包含所有樣本（策略 A）不僅反映真實篩檢情境（篩檢對象包含已有風險者），且排除與否對 AUC 的影響在可接受範圍內，佐證本研究結果的穩健性。

7.4 可解釋性分析

7.4.1 符號回歸的臨床意涵

符號回歸實驗發現了極簡但有效的預測公式：

1. **高血壓**：$0.130 \times \exp(SBP_{Y-2})$——僅使用四年前收縮壓，AUC 0.745，與 Random Forest（0.743）相近。公式極為簡潔，臨床可解釋性高。
2. **高血糖**：$0.114 \times FBG_{Y-1}$——簡單線性公式，僅使用兩年前空腹血糖，AUC 0.943，甚至略優於 LR（0.938），驗證了空腹血糖的高預測價值。
3. **高血脂**：$0.043 \times \exp(TC_{Y-2})$——使用四年前總膽固醇，AUC 0.801，低於 LR（0.867），較複雜的模型仍有提升空間。

7.4.2 簡單模型的實務潛力

單變數公式即具預測力（AUC 0.745-0.943），且公式符合臨床直覺——每項疾病的最佳預測因子皆為其診斷指標的歷史值。高血糖公式（AUC 0.943）甚至超越 LR，顯示簡單公式在某些任務中可達到最佳效能。這些簡單公式易於在基層醫療單位實施，不需複雜的計算資源，展現了可解釋 AI 在實務應用的潛力。

7.5 縱向資料的累積效益

健檢次數比較實驗顯示，累積更多次健檢紀錄有助於提升預測性能，但效益因疾病而異：

1. **高血壓受健檢次數影響最大**：從 1 次健檢（AUC 0.666）到 4 次健檢（AUC 0.835），提升幅度達 16.9 個百分點。特別是從 3 次到 4 次的跳躍最為顯著（0.676 → 0.835），顯示更長期的血壓縱向趨勢對高血壓預測具有關鍵價值。
2. **高血糖與高血脂受影響較小**：高血糖從 0.931 提升至 0.945（+1.4%），高血脂從 0.867 提升至 0.882（+1.5%）。這兩項疾病的單次健檢資料已具有高預測力，額外的縱向資料帶來的邊際效益有限。
3. **縱向資料的不對稱效益**：健檢次數的增加對不同疾病的預測改善幅度差異甚大，這與特徵重要性分析結果一致——高血壓預測較依賴多個時間點的血壓變化趨勢。

此結果為鼓勵民眾定期健檢提供了實證支持，尤其對高血壓風險族群而言，持續的縱向追蹤具有更大的預測價值。

7.6 多任務學習的適用性

MTL 未能展現預期優勢的可能原因包括：

1. **任務相關性偏低**：三項疾病的標籤相關性較弱（Phi 係數 < 0.1），共享特徵表示無法同時有效服務三項預測目標。
2. **資料規模充足**：滑動窗口資料集含 13,514 筆紀錄，各任務已有足夠樣本獨立訓練，MTL 的「資料增強」效益不顯著。

3. **任務難度不均**：高血糖 AUC 高達 0.93，高血壓僅 0.73，難度差距可能導致共享層的梯度被較易任務主導。

在本研究的資料集上，MTL 與 STL 的預測效能無顯著差異（AUC 差異 ≤0.8%）。考量到 STL 架構更為簡單且效能略優，獨立模型為合理選擇。

7.7 綜合討論

綜合以上各項實驗的討論，本研究的核心發現可歸納為以下幾點：

1. **模型效能**：Logistic Regression 在三項任務中皆表現穩定，簡單的線性模型在結構化健檢資料上已具備與複雜模型相當的排序能力。
2. **特徵重要性**：疾病特異性明顯，各疾病的最重要預測因子為其對應的診斷指標歷史值。Δ 特徵在 Top 10 中佔 30-50%，在資料受限情境下具有約 2% 的 AUC 提升效益。
3. **模型簡化**：Top 5 特徵即可達到接近完整效能，符號回歸的單變數公式亦具有實用的預測力。
4. **方法穩健性**：類別不平衡處理方法間無顯著差異，資料篩選策略的 AUC 變動 ≤ 1.3%，MTL 與 STL 差異 ≤0.8%，多項對比實驗皆驗證了研究方法的穩健性。
5. **縱向資料價值**：累積更多健檢紀錄有助於提升預測性能，尤其對高血壓預測的改善最為顯著（+16.9%），支持鼓勵民眾定期健檢的公共衛生政策。
6. **健康族群的預測行為**：對於歷次健檢指標皆正常且穩定的個體，模型將輸出低風險機率，此行為反映於各任務的高 Specificity 表現。更重要的是，Δ 特徵使模型能捕捉「數值仍在正常範圍內但已出現惡化趨勢」的亞健康狀態，搭配 SHAP 分析可指出個人化的主要風險因子，為預防醫學的早期介入提供依據。

第八章 結論與建議

8.1 研究總結

本研究旨在利用縱向健檢資料預測三高疾病（高血壓、高血糖、高血脂）的風險，並透過變化量特徵（Delta Features, Δ Features）的設計，捕捉個體健康狀態的動態變化，以提升預測性能與臨床應用價值。

本研究使用 Luo et al. (2024) 公開於 Dryad 的縱向健檢資料集（6,056 人），透過滑動窗口方法產生 13,514 筆建模紀錄，以 StratifiedGroupKFold 5-fold 交叉驗證進行系統性的模型比較與消融實驗。在十種機器學習模型的比較中，Logistic Regression 展現最穩定的表現（高血糖 AUC 0.938），且僅需 5 個關鍵特徵即可維持接近完整效能。Δ 特徵在資料受限情境下可帶來約 2% 的 AUC 提升，並在 Top 10 重要特徵中佔比達 30-50%，驗證了縱向特徵工程的價值。符號回歸發現的單變數公式（如高血糖 $0.114 \times FBG_{Y-1}$，AUC 0.943）進一步展現了簡單模型的實務潛力。

在學術層面，本研究將 Δ 特徵從單一疾病（Yang et al., 2025 的糖尿病預測）擴展至三高同時預測場景，並透過消融實驗提供跨疾病的實證依據。十種模型的系統性比較涵蓋傳統統計、樹模型、神經網路與符號回歸，搭配類別不平衡處理、資料篩選策略與多任務學習等多組對比實驗，從多個面向驗證了研究方法的穩健性。

在應用層面，SHAP 分析所識別的可干預風險因子（如血糖、血壓的變化趨勢）可輔助臨床判斷與衛教，而符號回歸的簡單公式與 Top 5 特徵模型則顯示精簡方案即可維持實用效能，有利於基層醫療單位部署早期預警系統。健檢次數與預測性能的正相關趨勢，亦為鼓勵民眾定期健檢提供了實證支持。

此外，本研究透過多組穩健性驗證實驗強化了上述結論的可信度：五種類別不平衡處理方法的 AUC 差異均小於 0.2%、資料篩選策略的影響不超過 1.3%、多任務學習與單任務學習的差異僅 0.3%，皆顯示主要發現不受特定實驗設定的左右，具有良好的方法穩健性。

8.2 研究限制

本研究存在以下限制：

資料限制

健檢族群的選擇偏差

本研究使用的縱向健檢資料來自自願參與健康檢查的族群，此族群相較於一般人口可能具有較高的健康意識與社經地位，因此研究結果的外推性（generalizability）可能受限。未來研究應納入更具代表性的一般人口樣本，或進行跨族群的外部驗證。

時間間隔不完全一致

由於健檢資料為真實世界資料，個體間的健檢時間間隔並非完全一致。雖然本研究採用滑動窗口方法盡可能標準化時間窗口，但時間間隔的變異仍可能影響 Δ 特徵的可比性。

缺乏飲食與生活型態資料

本研究僅使用生理檢驗指標作為預測特徵，缺乏飲食習慣、運動頻率、吸菸飲酒等生活型態資料。然而，美國膳食指南 2025-2030（Dietary Guidelines for Americans, DGA）明確指出，飲食模式與三高疾病具有直接且顯著的關聯。例如，添加糖攝取與代謝症候群、超加工食品與慢性疾病的發生皆有密切關係。若能整合飲食資料，預期可進一步提升預測性能並提供更全面的健康管理建議。

缺乏外部驗證資料集

由於缺乏具相似資料結構的外部資料集，本研究僅能透過內部交叉驗證評估模型性能。未來若能取得其他健檢中心或跨地區的縱向資料，將有助於驗證模型的穩健性與外推性。

模型限制

確診定義包含非生理指標條件

根據 Luo et al. (2024) 原始論文，三項疾病的確診標記並非皆以生理指標閾值為唯一依據：高血壓定義為 $SBP \geq 140$ 或 $DBP \geq 90$ mmHg，**或已確診且正在服用降壓藥物**；

高血糖定義為 FBG ≥ 7.0 mmol/L，**或自我報告糖尿病**；高血脂則僅以 TC ≥ 6.22 mmol/L 為標準，無額外條件。

此定義差異對高血壓與高血糖的預測模型產生影響：正在服用降壓藥的個體，其血壓可能已被控制於正常範圍，但仍被標記為患病，導致特徵與標籤之間出現不一致。此現象亦可能使 Δ 特徵產生誤導——例如，開始服藥的個體可能呈現 ΔSBP < 0（血壓下降），模型將此解讀為低風險訊號，但實際上該個體已確診高血壓。類似地，自我報告糖尿病但 FBG 受控的個體亦存在同樣問題。高血脂因僅使用生理閾值定義，不受此限制影響。

雖然本研究的排除策略 C（§6.6）已排除 Y-2 時生理指標超標的既有患者，但無法識別「Y-2 時已用藥且指標被控制於正常範圍」的個體，此為使用公開資料集的固有限制。未來研究若能取得用藥紀錄，應納入用藥史作為額外特徵或排除條件，以更精確地定義新發個案。

未探索深度時間序列模型

本研究因時間點數量有限（3 個時間點），未能充分探索長短期記憶網路（Long Short-Term Memory, LSTM）、門控循環單元（Gated Recurrent Unit, GRU）等深度時間序列模型的潛力。這些模型在處理更長序列的時間序列資料時可能展現更優異的性能。

8.3 未來研究方向

基於本研究的發現與限制，建議未來研究可朝以下方向發展：

方法論改進

深化多任務學習

本研究的 MLP 多任務學習實驗顯示 MTL 與 STL 無顯著差異（AUC 差異 ≤ 0.8%），可能因三項疾病間的標籤相關性偏低，且任務難度不均導致共享層梯度被較易任務主導。未來可探索動態權重調整（dynamic weight tuning）或梯度平衡（gradient balancing）等技術，以改善多任務學習在類別不平衡與任務難度不均場景下的表現。

可解釋性分析深化

雖然本研究透過 SHAP 與符號回歸進行了初步的可解釋性分析，但仍可進一步探索個體層級的預測解釋（individual-level explanations）與反事實解釋（counterfactual explanations），以支持更精細的個人化健康建議。

納入飲食與生活型態特徵

如前述研究限制所提及，本研究僅使用生理檢驗指標，缺乏飲食與生活型態資料。然而，DGA 明確指出飲食模式（添加糖攝取、超加工食品頻率、脂肪品質等）與三高疾病具有直接關聯，且飲食介入可預防或減緩疾病進展。若未來研究能取得飲食資料，建議納入超加工食品攝取頻率、添加糖攝取量、飽和脂肪比例等特徵。相較於年齡、性別等不可改變的因子，飲食習慣是高度可干預的風險因子，結合生理指標與飲食特徵預期可同時提升預測性能與健康管理的可操作性。

模型與資料擴展

時間序列深度學習

若未來能取得更長時間序列的健檢資料（如 5 個以上時間點），建議探索 LSTM、GRU 或轉換器模型（Transformer）等深度時間序列模型，以充分利用縱向資料的時序資訊。

外部驗證與跨族群研究

建議與其他健檢中心或研究機構合作，取得外部驗證資料集，評估模型在不同族群、地區的穩健性與外推性。此外，也可探索模型在不同年齡層、性別、疾病嚴重程度的表現差異。

擴展至其他慢性疾病

本研究的 Δ 特徵工程方法與模型比較框架可推廣至其他慢性疾病預測，如心血管疾病、腎臟疾病、代謝症候群等，以驗證方法的通用性。

納入多模態資料

未來可整合更多元的資料來源，如用藥史、家族病史、基因資料、穿戴式裝置的活動數據等，建立更全面的健康風險評估模型。

參考文獻

英文文獻

Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. F., & van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. PLoS ONE, 14(5), e0213653. https://doi.org/10.1371/journal.pone.0213653

Alberti, K. G. M. M., Eckel, R. H., Grundy, S. M., Zimmet, P. Z., Cleeman, J. I., Donato, K. A., Fruchart, J.-C., James, W. P. T., Loria, C. M., & Smith, S. C., Jr. (2009). Harmonizing the metabolic syndrome: A joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation, 120(16), 1640–1645. https://doi.org/10.1161/CIRCULATIONAHA.109.192644

American Diabetes Association. (2025). Standards of care in diabetes—2025. Diabetes Care, 48(Suppl 1).

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression.jl. arXiv. https://doi.org/10.48550/arXiv.2305.01582

Dinh, A., Miertschin, S., Young, A., & Mohanty, S. D. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Medical Informatics and Decision Making, 19(1), 211. https://doi.org/10.1186/s12911-019-0918-5

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hung, M.-H., Shih, L.-C., Wang, Y.-C., Leu, H.-B., Huang, P.-H., Wu, T.-C., Lin, S.-J., Pan, W.-H., Chen, J.-W., & Huang, C.-C. (2021). Prediction of masked hypertension and masked uncontrolled hypertension using machine learning. Frontiers in Cardiovascular Medicine, 8, 778306. https://doi.org/10.3389/fcvm.2021.778306

James, P. A., Oparil, S., Carter, B. L., Cushman, W. C., Dennison-Himmelfarb, C., Handler, J., Lackland, D. T., LeFevre, M. L., MacKenzie, T. D., Ogedegbe, O., Smith, S. C., Jr., Svetkey, L. P., Taler, S. J., Townsend, R. R., Wright, J. T., Jr., Narva, A. S., & Ortiz, E. (2014). 2014 evidence-based guideline for the management of high blood pressure in adults: Report from the panel members appointed to the Eighth Joint National Committee (JNC 8). JAMA, 311(5), 507–520. https://doi.org/10.1001/jama.2013.284427

Kanegae, H., Suzuki, K., Fukatani, K., Ito, T., Harada, N., & Kario, K. (2020). Highly precise risk prediction model for new-onset hypertension using artificial intelligence techniques. The Journal of Clinical Hypertension, 22(3), 445–450. https://doi.org/10.1111/jch.13759

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems 30 (pp. 3146–3154).

Liu, Y.-Q., Chang, T.-W., Lee, L.-C., Chen, C.-Y., Hsu, P.-S., Tsan, Y.-T., Yang, C.-T., & Chu, W.-M. (2024). Use of machine learning to predict the incidence of type 2 diabetes among relatively healthy adults: A 10-year longitudinal study in Taiwan. Diagnostics, 15(1), 72. https://doi.org/10.3390/diagnostics15010072

Luo, Y., Wu, Q., Meng, R., Lian, F., Jiang, C., Hu, M., Wang, Y., & Ma, H. (2023). Associations of serum uric acid with cardiovascular disease risk factors [Dataset]. Dryad Digital Repository. https://doi.org/10.5061/dryad.z08kprrk1

Luo, Y., Wu, Q., Meng, R., Lian, F., Jiang, C., Hu, M., Wang, Y., & Ma, H. (2024). Associations of serum uric acid with cardiovascular disease risk factors: A retrospective cohort study in southeastern China. BMJ Open, 13(9), e073930. https://doi.org/10.1136/bmjopen-2023-073930

Majcherek, D., Ciesielski, A., & Sobczak, P. (2025). AI-driven analysis of diabetes risk determinants in U.S. adults: Exploring disease prevalence and health factors. PLoS ONE, 20(9), e0328655. https://doi.org/10.1371/journal.pone.0328655

National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (2002). Third report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III) final report. Circulation, 106(25), 3143–3421. https://doi.org/10.1161/01.CIR.0000038419.01177.FA

Ohira, T., & Iso, H. (2013). Cardiovascular disease epidemiology in Asia: An overview. Circulation Journal, 77(7), 1646–1652. https://doi.org/10.1253/circj.CJ-13-0702

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. https://doi.org/10.1038/323533a0

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE, 10(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Stanciu, S., Rusu, E., Miricescu, D., Radu, A. C., Axinia, B., Vrabie, A. M., Ionescu, R., Jinga, M., & Sirbu, C. A. (2023). Links between metabolic syndrome and hypertension: The relationship with the current antidiabetic drugs. Metabolites, 13(1), 87. https://doi.org/10.3390/metabo13010087

Sun, D., Liu, J., Xiao, L., Liu, Y., Wang, Z., Li, C., Jin, Y., Zhao, Q., & Wen, S. (2017). Recent development of risk-prediction models for incident hypertension: An updated

systematic review. PLoS ONE, 12(10), e0187240. https://doi.org/10.1371/journal.pone.0187240

Sun, Z., & Zheng, Y. (2025). Metabolic diseases in the East Asian populations. Nature Reviews Gastroenterology & Hepatology, 22(7), 500–516. https://doi.org/10.1038/s41575-025-01058-8

Tsai, H., Yang, T.-W., Wu, T.-Y., Tu, Y.-C., Chen, C.-L., & Chou, C.-F. (2025). Multitask learning multimodal network for chronic disease prediction. Scientific Reports, 15(1), 15468. https://doi.org/10.1038/s41598-025-99554-z

U.S. Department of Agriculture and U.S. Department of Health and Human Services. (2025). Dietary Guidelines for Americans, 2025–2030 (10th ed.). https://www.dietaryguidelines.gov/

Wang, C.-C., Chu, T.-W., & Jang, J.-S. R. (2024). Next-visit prediction and prevention of hypertension using large-scale routine health checkup data. PLoS ONE, 19(11), e0313658. https://doi.org/10.1371/journal.pone.0313658

World Health Organization. (2023). Global report on hypertension: The race against a silent killer. WHO. https://www.who.int/publications/i/item/9789240081348

World Heart Federation. (2023). World Heart Report 2023: Confronting the world's number one killer. WHF. https://world-heart-federation.org/world-heart-report-2023/

Yang, C.-C., Wu, S.-T., Chu, T.-W., Liu, C.-H., & Chuang, Y.-J. (2025). Dual machine learning framework for predicting long-term glycemic change and prediabetes risk in young Taiwanese men. Diagnostics, 15(19), 2507. https://doi.org/10.3390/diagnostics15192507

Ye, C., Fu, T., Hao, S., Zhang, Y., Wang, O., Jin, B., Xia, M., Liu, M., Zhou, X., Wu, Q., Guo, Y., Zhu, C., Li, Y., Culver, D. S., Alfreds, S. T., Stearns, F., Sylvester, K. G., Widen, E., McElhinney, D., & Ling, X. (2018). Prediction of incident hypertension within the next year: Prospective study using statewide electronic health records and machine learning. Journal of Medical Internet Research, 20(1), e22. https://doi.org/10.2196/jmir.9268

Zhao, D. (2021). Epidemiological features of cardiovascular disease in Asia. JACC: Asia, 1(1), 1–13. https://doi.org/10.1016/j.jacasi.2021.04.007

中文文獻

國民健康署（2022）。*2017–2020 國民營養健康狀況變遷調查*。衛生福利部。https://www.hpa.gov.tw/Pages/Detail.aspx?nodeid=3999&pid=15562

國民健康署（2025）。*成人預防保健服務擴大至 30 歲公告*。衛生福利部。https://www.hpa.gov.tw/Pages/detail.aspx?nodeid=4878&pid=18755