

出國報告（出國類別：進修）

利物浦大學
心臟血管醫學中心
進修心得

服務機關：醫學研究部/基礎醫學科/生物統計小組

姓名職稱：陳韻仔 契約副研究員

派赴國家/地區：英國/利物浦

出國期間：113 年 08 月 31 日起至 113 年 10 月 01 日

報告日期：113 年 10 月 08 日

摘要

感謝臺中榮民總醫院與財團法人榮康醫學發展基金會的支持，讓我於 2024 年 9 月 1 日至 9 月 30 日，得以前往英國利物浦大學心臟血管醫學中心進行為期一個月的進修。此次進修旨在促進雙方的合作與交流，並針對心血管疾病的預測模型進行開發與研究，進一步提升研究能力。

在這期間，我與英國利物浦大學心臟血管醫學中心 Pro. Gregory Lip 教授及其博士生團隊、Prof. Ya-Lin Zheng 教授的 AI 團隊密切合作，進行了時間序列與大數據的人工智慧處理，針對金山心血管世代研究數據進行 Lipid Profile 的長期變化趨勢分析，並開發機器學習預測與深度學習的風險預測模型。此外，合作探討了 AF 電燒術後復發及二次電燒的心電圖預測模型，並學習使用 Pythob 處理心電圖圖像轉數據並去識別化的方法。此外，我們利用 TriNetX 數據資料庫，對 SGLT2i 藥物與中風療效進行比較分析，進一步探討了不同藥物之間的效果差異，為未來的臨床應用提供了新的見解。

此次進修不僅加深了我對心血管疾病與人工智慧應用的理解，也為未來的國際合作打下了良好的基礎。

關鍵字：心血管疾病預測模型、人工智慧應用、深度學習、心電圖分析、臨床數據分析

目 次

一、 目的	8
二、 過程	8
三、 心得	20
四、 建議事項.....	22
(一) 整合多次時序變化資料以提升預測模型的準確性	
(二) 建置網路雲端平台以促進 AI 模型的應用與共享	
(三) 建議心血管研究加強心電圖數據的收集與標準化處理常規化	
(四) 積極爭取國際數據進行人工智慧模型外部驗證	
五、 附錄	23

一、 目的

此次前往英國利物浦大學心臟血管醫學中心進行為期一個月的進修，旨在促進臺中榮民總醫院與國際頂尖心血管研究機構的學術交流與合作，進一步提升我在心血管疾病研究及人工智慧技術應用領域的專業能力。本次進修的主要目的是學習最新的時間序列分析與大數據處理技術，並運用深度學習模型（如 Recurrent Neural Network, RNN）開發精準的心血管疾病風險預測模型。

此外，此行亦著重於強化心電圖數據的收集與標準化處理技術，提升對心房顫動（AF）電燒術後復發及二次電燒需求的預測準確性。同時，我期望通過此次進修，能夠與國際學者共同探討如何應用人工智慧技術改善臨床決策，並與國際醫療機構建立長期合作，爭取外部數據驗證的機會，以提高模型在不同族群中的通用性和應用價值。這些學習成果將有助於我將先進技術引入臺灣，進一步推動本院在心血管疾病預測與預防領域的發展。

二、 過程

地點：英國利物浦大學

時間：2024/9/1-2024/9/30

進修人員：陳韻仔博士（臺中榮民總醫院/醫學研究部/副研究員 兼 國立中興大學/學士後醫學系/助理教授）

合作研究人員及單位：

Pro. Gregory Lip, Price-Evans Chair of Cardiovascular Medicine,
University of Liverpool

Dr. Yang Chen, PhD student of Pro. Lip, University of Liverpool

Dr. Huang Bi, PhD student of Pro. Lip, University of Liverpool

Prof. Ya-Lin Zheng, Professor of AI in Healthcare, Department of Eye and
Vision Sciences, Institute of Life Course and Medical Sciences

Peter Thompson, Student of Prof. Ya-Lin Zheng, University of Liverpool

Ruo-Jun Zhang, Student of Prof. Ya-Lin Zheng, University of Liverpool

Dr. Chih-Min Liu, Taipei Veterans General Hospital

2024 年 9 月研究進度總結

在 2024 年 9 月期間，主要進行了以下幾項研究與合作活動：

1. 時間序列與大數據人工智慧處理分析：

與 Pro. Gregory Lip、Prof. Ya-Lin Zheng 及其博士學生 Dr. Yang Cheng、Dr. Huang Bi、Ruo-Jun Zheng、Peter Thompson 合作

- 與 Pro. Lip 的學生 Dr. Yang Chen 合作，針對金山心血管世代研究的多次 Lipid Profile（例如 Apo A1、Apo B、LDL、HDL、Triglycerides）的隨時間變化趨勢和變異度（三次定期隨訪資料→Baseline:1990-1991、世代 2: 1994-1995、世代 5: 2006），研究其對冠心病、中風及各類死亡原因的群體趨勢影響。
- 與 Pro. Lip 的學生 Dr. Yang Chen 和 Dr. Huang Bi 共同學習 machine learning，使用波蘭大數據進行機器學習模型的開發和預測分析。
- 討論 AF 電燒術後復發及需進行第二次電燒的預測模型，計畫與臺北榮總劉志民醫師合作，利用心電圖數據進行模型開發：在 Prof. Ya-Lin Zheng 及其學生 Peter Thompson 的指導下，學習如何將 ECG（心電圖）pdf 圖像轉換為數據格式，並進行去辨識化處理，以便進一步分析和建模。
- 預計使用 Deep Learning 技術（Recurrent Neural Network, RNN），並嘗試預測未來疾病發生的可能性：和 Pro. Ya-Lin Zheng 的學生 Ruo-Jun Zhang 及 Pro. Lip 的學生 Yang Chen，處理 CCCC 的時間序列數據，使用 Deep learning model（Recurrent Neural Network, RNN）做 Time-dependent ROC，不同存活時間對後續多段終點預測模型。
- 修改並準備投稿 CCCC SCD 預測 AI 模型。

2. 公用巨量數據研究討論：

與 Pro. Gregory Lip 其博士學生 Dr. Yang 合作

- 整理 Taiwan Biobank 在臺灣健康加值中心串聯預後的 Life Essential 8 資料，並預期與 UK Biobank 數據進行對比分析。
- TriNetX 數據分析合作：臺中榮總與 Liverpool University 分別都各自可以申請分析權限→

與 Dr. Yang Chen 共同分析了兩項研究，涉及中風與 SGLT2i 藥物的效果比較：

- Study 1: DM+Stroke without AF：納入年齡 20 至 90 歲患者（排除 55 至 85 歲），使用 ICD-10-CM I63 全部代碼，分析無 AF 患者的中風情況。
- Study 2: DM+Stroke：納入同樣年齡段（20 至 90 歲）中風住院患者，不排除 AF，進行 SGLT2i 與非 SGLT2i 藥物效果的比較。
- 在這兩項研究中，進一步比較了 Empagliflozin、Canagliflozin 和 Dapagliflozin 三種藥物之間的效果差異。
- 主要觀察的 outcome 包括：AF、心衰竭、心室心律不整(VT)、Takotsubo、慢性腎病、失智症、憂鬱症、非致命性心肌梗塞、非致命性缺血性中風（再中風）、全因死亡率和心血管死亡率。

2024. 09. 01

抵達 Gatwick Airport，轉換到 Liverpool University 附近民宿。

2024. 09. 02

和 Pro. Lip (Price-Evans Chair of Cardiovascular Medicine, at the University of Liverpool, UK - and Director of the Liverpool Centre for Cardiovascular Science at the University of Liverpool and Liverpool Heart & Chest Hospital) 的博士學生 Dr. Yang, Chen，共同討論合作研究可行性。



2024. 09. 03

09:00 AM 和 Pro. Lip 及 Dr. Yang Chen，共同討論未來的研究方向：

1. 討論簽訂保密協議：討論研究資料不外流，在自己醫院分析處理。
2. 與北榮劉志民醫師共同合作，做第一次 AF 電燒之後，是否會發生 AF 復發，

以及需要第二次電燒的心電圖預測模型。

3. 與 Dr. Yang Chen 合作，做金山心血管世代研究的多次 lipid profile (e.g. Apo A1/ Apo B/LDL/HDL/Triglycerides) 隨時間的多次變化趨勢和變異度，對冠心病、中風、各種死因的影響趨勢。
4. 把 CCCC SCD prediction AI model 改格式投稿。
5. 整理 Taiwan Biobank 的 Life Essential 8 資料，和 UK Biobank 做比較。

10:30 AM 和 Pro. Ya-Lin Zheng (Professor of AI in Healthcare, Department of Eye and Vision Sciences | Institute of Life Course and Medical Sciences) 及 Dr. Yang Chen，共同討論未來的研究方向：

與 Pro. Ya-Lin Zheng 研究室學習：

- a. 處理 ECG pdf 影像，成為數據資料，並去辨識化。
- b. 處理 CCCC 時間變化追蹤資料，使用 Deeping learning，預測未來發生疾病的可能性。

18:00 PM Pro. Lip, 研究團隊聚餐

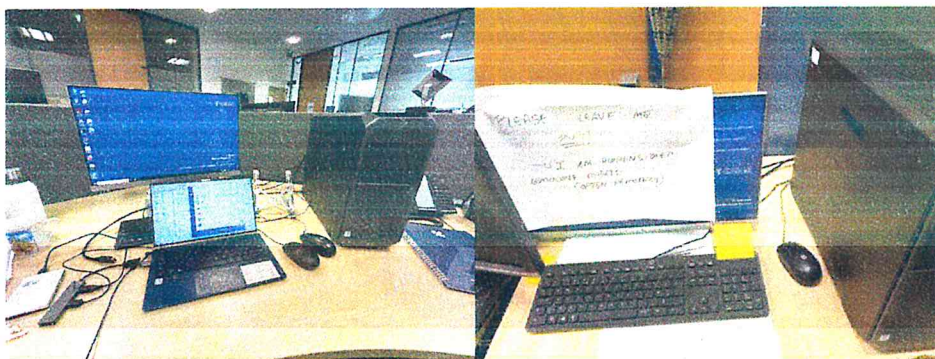


2024. 09. 09

1. 與 Dr. Yang Chen 和 Dr. Huang Bi 共同學習 machine learning，使用波蘭大數據進行 ML 預測模型 (python)。
2. 把 CCCC SCD prediction AI model 改方法，和 Pro. Lip 確認後再投稿。

2020. 09. 11

1. 分一台配有 GPU 的超級電腦給我，方便處理人工智慧模型和心電圖影像處理。與 Prof. Ya-Lin Zheng 的學生 Peter Thompson 約 4 PM，學習如何將心電圖圖像檔 (pdf) 轉成國際數據格式，並進行匿名化處理。



2. 今日與 Dr. Yang Chen 共同進行 TriNetX 數據分析，針對兩個研究進行以下調整與討論：

Study 1: Stroke without AF

納入條件：年齡範圍 20 至 90 歲（排除 55 至 85 歲）。

Stroke 的 ICD 10 CM：將條件修正為使用 I63 的所有代碼。

AF 的排除條件保持不變。

Study 2: Stroke（不排除 AF，包含所有中風住院患者）

納入條件：年齡範圍 20 至 90 歲（排除 55 至 85 歲）。

Stroke 的 ICD 10 CM：同樣使用 I63 的所有代碼。

針對以上兩個研究，我們進行以下比較：

- a. 比較 SGLT2i 與非 SGLT2i 藥物的效果。
- b. 進行 Empagliflozin、Canagliflozin、Dapagliflozin 三者之間的差異比較，進行兩兩對比。

Outcome 指標：每個 outcomes 分開檢視，主要包括：

Atrial fibrillation (new-onset AF)、heart failure、severe ventricular arrhythmias: VT (只要 VT)、Takotsubo、chronic kidney disease、dementia、depression、non-fatal myocardial infarction、non-fatal ischemic stroke (re-stroke)

All-cause death、cardiovascular death/mortality

2020. 09. 12

16:30 PM

和 Pro. Ya-Lin Zheng 的學生 Peter 學習處理 ECG pdf 影像，成為數據資料，並去辨識化（參考附錄一）。

2020. 09. 13

10:30 AM 和 Pro. Ya-Lin Zheng 的學生 Ruo-Jun Zhang 習處理 CCCC 時間變化追蹤資料，使用 Deeping learning，預測未來發生疾病的可能性，以及分析到發病的時間。討論使用 deep learning model (Recurrent Neural Network, RNN) 做 Time-dependent ROC model，採用存活到某時間點（多段）對往後（不同時間點）存活率的預測和多死因死亡的預測準確度，以及計算單一指標的累積發生風險。

Ref:

1. Lee, Changhee, Jinsung Yoon, and Mihaela Van Der Schaar. "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data." *IEEE Transactions on Biomedical Engineering* 67.1 (2019): 122-133.
2. Hochreiter, S. "Long Short-term Memory." *Neural Computation MIT-Press* (1997).

2020.09.16

09:00 AM 與 Dr. Yang Chen 和 Dr. Huang Bi 共同學習如何使用 R 和 Python，操作基本 machine learning，使用波蘭大數據進行 CABG 後會發生 Atrial fibrillation 的預測模型。



11:00 AM 和 Pro. Ya-Lin Zheng 的學生 Ruo-Jun Zhang 學習處理 CCCC 時間變化追蹤資料，使用 Deeping learning，預測未來發生疾病的可能性，以及分析到發病的時間，建立單一指標不同連續數值的累積風險發生率。

19:00 PM Pro. Ya-Lin Zheng 和 Pro. Lip 中秋聯合晚宴。



2020.09.17

參加人工智慧研究成果討論會

[CONFIRMED VENUE: LJMU STUDENT LIFE BUILDING ROOM 203]

Agenda:

10.00 - Welcome

10.10 - Dr Ryan Bellfield

10.30 - Dr Bi Huang

10.50 - Dr Dharmesh Mistry 急診 AF 分型與預後的預測模型

11.10 - Dr Peter Thompson

11.30 - George Margereson

11.50 - Discussion



2020.09.18

09:00 AM On-line research meeting with Taiwan team:

Meeting summary→

Package 1 involves participants from the baseline data (TWB2 cohort).

For Article 1, the primary predictors include LE7 (the main variable) along with each domain in lifestyle factors.

Additionally, genetic markers such as PGS000035 (for atrial fibrillation) and PGS000911 (for ischemic stroke in AF cases) are being considered. The outcomes being studied are atrial fibrillation (AF) and AF-related strokes.

This article will explore the interaction between lifestyle factors and genetic components, with the AUC (Area Under the Curve) used to assess life essential (LE) scores combined with genetic data.

>>> for WP1, will use the whole Baseline cohort (n=35k)

Paper 1 is the baseline characteristics of LE7, and the prediction of AF and AF related complications

Paper 2: Polygenic risk Score versus Polygenenic risk score plus LE7 in prediction of incident AF and AF-complications (stroke)

Analysis: AUC, NRI, IDI

Package 2 focuses on participants from the first follow-up.

For Article 2, the main predictors are LE8 (the primary variable), LE7, and other lifestyle domains, as well as genetic data from PGS000039 (for ischemic stroke). The key outcomes are stroke and death.

Similar to Package 1, we are examining the interaction between lifestyle factors and genetic influences.

For Article 3, the focus is on analyzing the delta changes in LE7 (lifestyle scores) between the baseline and the first follow-up, utilizing these changes as a predictor. Specifically, the analysis employs a Markov model to understand the dynamic transitions between lifestyle states and their impact on health outcomes.

For WP2 will use the 1st Followup TWB cohort (n=18k)

Paper 1 is LE8, in prediction of incident stroke and AF-related stroke

Paper 2 is LE8 in prediction of death and CV death

Paper 3 is comparing change in LE7 in the 18k between baseline and 1st followup ('delta LE7') in prediction of AF, stroke, death

Paper 4 is polygenic risk scores plus LE8 (versus polygenic risk scores alone) in prediction of AF, stroke, death

2020.09.19

19:00 PM Pro. Pro. Lip 團隊聚餐。



2020.09.23

15:30 PM 和 Pro. Ya-Lin Zheng 的學生 Ruo-Jun Zhang 及 Pro. Lip 的學生 Yang Chen 確認使用 Deep learning model (Recurrent Neural Network, RNN) 做 Time-dependent ROC, 不同存活時間對後續多段終點的結果, 預測和多死因死亡的預測準確度, 以及計算單一指標的累積發生風險。

Ref:

Changhee Lee, et al. IEEE Trans Biomed Eng. 2020 Jan;67(1):122-133

Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data

方法: 使用 Dynamic Deephit, 利用 CCCC 共 1108 人三次的測量數據: 包括 HDL、LDL、APOA1、APOB、fasting glucose, 調整 age, sex, HTN, DM, hyperlipidemia, ckd, cad 等變數, 採用 80% 當作訓練集, 20% 當作測試集。

2024.09.25

和 Pro. Lip 及 Dr. Yang Chen，共同討論研究結果撰寫成正式稿件：

針對金山心血管世代研究的多次 Lipid Profile (例如 Apo A1) 的隨時間變化趨勢和變異度 (三次定期隨訪資料→Baseline:1990-1991、世代 2: 1994-1995、世代 5: 2006)，研究其對冠心病、中風及各類死亡原因的群體趨勢影響，其已完成文摘如下，預計投稿 2025 年歐洲心臟年會 (ESC 2025) 並盡快完成論文進行投稿。研究文摘，參考附錄二。

2020.09.27

15:30 PM 和 Pro. Ya-Lin Zheng 的學生 Ruo-Jun Zhang 及 Pro. Lip 的學生 Yang Chen 確認使用 Deep learning model (Recurrent Neural Network, RNN) 做 Time-dependent ROC，不同存活時間對後續多段終點的討論修正結果，方法及初步結果，參考附錄三。

2024.09.30 - 2024.10.01

從 Liverpool 搭乘火車轉換到 London，再抵達 Gatwick Airport，搭乘飛機回臺北。

三、心得

臺中榮民總醫院醫學研究部與英國利物浦大學心臟血管醫學中心簽訂契約，建立共同合作關係，建立學習交流計劃，讓研究人員至利物浦大學心臟血管醫學中心進修學習，提升研究能力，致力發展共同研究計畫擬開發並完成心血管疾病的預測模型（如附件）。2024年9月1日至9月30日期間，感謝臺中榮民總醫院給我機會，前往英國利物浦大學進行為期一個月的進修，也誠摯感謝財團法人榮康醫學發展基金會此次對於我出國進修期間的贊助。本次進修不僅是我專業成長中的重要里程碑，更為我在心血管疾病研究與人工智慧技術的交叉應用提供了寶貴的學習與研究機會。此次訪問使我能夠與國際頂尖學者深入交流，並與多方專業團隊共同探討臨床研究的創新方法，進一步強化了我在該領域的研究基礎與實踐能力。

本次進修的主題涵蓋時間序列分析、大數據處理及人工智慧模型的臨床應用。這些領域不僅是當前醫學研究的前沿課題，亦是臨床決策支持系統中極具潛力的技術應用。我非常榮幸能夠與利物浦大學心臟血管醫學中心的 Pro. Gregory Lip 教授和 AI 團隊的 Prof. Ya-Lin Zheng 教授及其聯合團隊密切合作，參與多項涵蓋心血管疾病、預測模型開發以及生物標記物長期變化分析的跨學科研究。

在心血管疾病的長期趨勢分析方面，我與 Pro. Gregory Lip 教授的博士生 Dr. Yang Chen 及其團隊合作，運用金山社區心血管世代研究的數據，對多次 Lipid Profile（例如 Apo A1、Apo B、LDL、HDL、Triglycerides）的隨時間變化趨勢進行分析，並探討這些變化與死亡、冠心病及中風等重大不良心血管事件（MACE）之間的關聯。我們透過 Group-based multivariate trajectory modeling 將參與者分類為不同的脂質變化軌跡群，並評估其與長期心血管事件風險之間的關聯性。這項研究強調了動態監測脂質水平的重要性，並為高風險族群的早期風險分層提供了新的視角。

此外，我也參與了以人工智慧與深度學習技術進行的心血管疾病預測模型開發工作。我與 Prof. Ya-Lin Zheng 教授及其博士生合作，應用深度學習模型（Recurrent Neural Network, RNN），處理來自 CCCC 的時間序列數據，並進行 Time-dependent ROC 分析。我們通過此模型預測了不同生存時間點的心血管疾病發生風險，並且驗證了模型在多段不同時間與不同終點預測中的精確性。

在數據處理技術方面，我參與了將心電圖（ECG）影像轉換為數據格式並去辨識化的項目，這是一項對大規模臨床數據進行標準化處理的重要步驟。我與團隊密切合作，學

習並實踐如何將 ECG PDF 影像進行預處理，為後續進行心房顫動（AF）電燒術後復發的預測模型開發打下基礎。此次進修的另一重要成果是與臺北榮總劉志民醫師共同合作，探討首次 AF 電燒術後患者的復發風險，以及預測是否需要進行二次電燒的模型開發。

最後，因臺中榮民總醫院和 Liverpool University 都各自有權限可以進行 TriNetX 的研究，我與 Dr. Yang Chen 博士候選人共同討論利用 TriNetX 可以進行的研究主題，評估其研究新穎與可行性。我們探討了 SGLT2i 類藥物對於不同中風患者的療效比較。在可行的兩項研究中，我們嘗試分析了 SGLT2i 與非 SGLT2i 藥物對於合併或未合併心房顫動的糖尿病患者的療效在腦神經預後的差異（包括：中風與憂鬱症），並深入探討了三種 SGLT2i 藥物（Empagliflozin、Canagliflozin 和 Dapagliflozin）的效果比較，對未來臨床應用具有參考價值。

總結來說，此次英國利物浦大學的進修經歷對我來說是一段極具啟發性與挑戰性的旅程。透過與多方專家的密切合作，我不僅在專業技能上獲得了顯著提升，亦開啟了多個國際合作研究的契機，比如說可能取得歐洲研究聯盟（如：波蘭急重症資料數據）的研究數據做外部驗證。此外，我也將本身傳統流行病學與預防醫學和收案的專業，指導 Liverpool University 的 AI 研究團隊，將傳統流病背景也應用於 AI 的模型。希望此次的合作與學習經驗，可以應用於臺灣的臨床研究，特別是在心血管疾病的預測與風險管理方面，為推動更精準的醫療決策提供助力，我深信這次經驗將為我未來的研究與臨床工作帶來長遠影響。

四、建議事項

甲、整合多次時序變化資料以提升預測模型的準確性：

心血管疾病的預測模型應充分考慮來自多次隨訪的時序變化資料，像血液生化檢查如 Lipid Profile，或是超音波檢查、心電圖等長期變化數據的動態分析。這將有助於更準確地反映不同時間的數據變化對於疾病風險和競爭風險趨勢的影響，可以提升模型的預測準確性。建議將數據分階段處理，並根據不同時間點的變化趨勢調整模型參數，使其更符合臨床實際。

乙、建置網路雲端平台以促進 AI 模型的應用與共享：

為了有效推動 AI 模型在臨床與研究中的應用，本院可考慮設置專屬的網路雲端平台，供研究人員架設已開發之 AI 預測模型。該平台可提供期刊審稿人、研究人員及臨床醫師查閱已開發模型使用，甚至可以開放部分功能供民眾進行風險自我評估，促進 AI 技術的臨床應用。

丙、建議心血管研究加強心電圖數據的收集與標準化處理常規化：

心電圖數據對心血管疾病的預測至關重要，特別是在心律不整和心因性猝死的預測模型中。建議將心電圖數據的收集和標準化處理常規化，以避免後續處理時基準不一致可能造成的偏差。可採取的改進作法包括引入統一的數據收集格式與自動化處理系統，確保所有數據的質量與一致性，進而提高預測模型的準確性與臨床應用價值。

丁、積極爭取國際數據進行人工智慧模型外部驗證外部驗證：

由於亞洲和歐美族群的臨床數據可能存在顯著差異，預測模型的外部驗證至關重要。建議持續爭取與國外醫療機構及研究中心的合作，進行多中心、多族群的數據比較和驗證。這不僅能提升模型的可信度和通用性，還有助於確認模型在不同地區和族群的適用性，推動國際學術交流。利用本院、本國多中心、與國外重要數據進行內外部驗證，並將驗證結果納入期刊發表與臨床應用指導中，以提升 AI 模型的國際影響力。

五、 附錄

附錄一

和 Pro. Ya-Lin Zheng 的學生 Peter 學習處理 ECG pdf 影像，成為數據資料，並去辨識化：

參考：Python Code: raster_pdf_to_csv (python file)

1、 將心電圖 pdf 檔案和 raster_pdf_to_csv (python file)放在同一個資料匣

2、 安裝 Python，使用 Command 命令提示字元→

```
pip install numpy
pip install pymupdf
pip install h5py
pip install matplotlib
```

3、 使用 command 命令提示字元→

更改檔案位置到

```
cd D:\副研究員\Lip\20240903\ECG python\extracted
```

```
D:
```

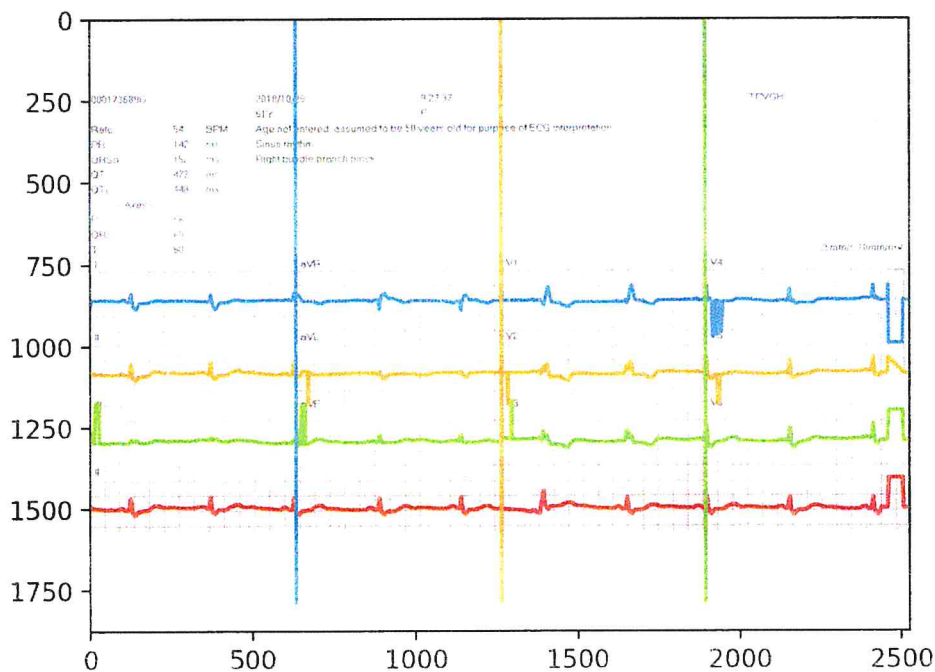
```
C:\Users\YYChen>cd D:\副研究員\Lip\20240903\ECG python\extracted
```

```
C:\Users\YYChen>D:
```

```
D:\副研究員\Lip\20240903\ECG python\extracted>
```

```
D:\副研究員\Lip\20240903\ECG python\extracted>python raster_pdf_to_csv.py
"GP2_PRE_PDF(28)\000173689G_20181029.pdf"
```

可以將原始單一 pdf 檔案讀取成以下畫面→



跳到一次放很多 pdf 的資料匣

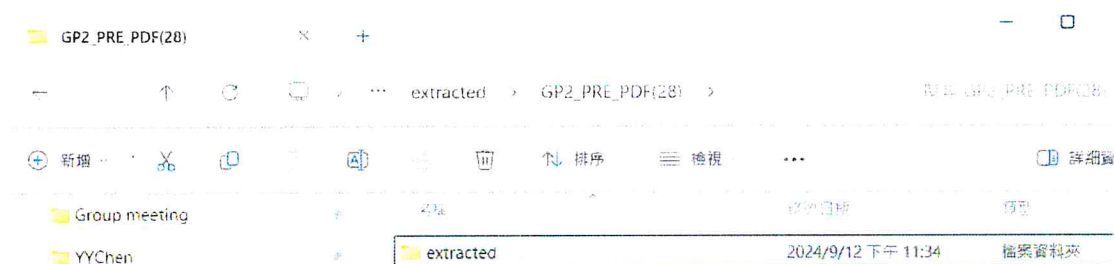
D:\副研究員\Lip\20240903\ECG python\extracted>cd "GP2_PRE_PDF(28)"

```
C:\Users\YYChen>cd D:\副研究員\Lip\20240903\ECG python\extracted
C:\Users\YYChen>D:
D:\副研究員\Lip\20240903\ECG python\extracted>cd "GP2_PRE_PDF(28)"
D:\副研究員\Lip\20240903\ECG python\extracted\GP2_PRE_PDF(28)>
```

D:\副研究員\Lip\20240903\ECG python\extracted>cd "GP3_PRE_PDF(53)"

再建立一個要放所有輸出的資料匣"extracted"，跑出換成 photo 後要轉成 pdf 的資料匣位置→

D:\副研究員\Lip\20240903\ECG python\extracted\GP2_PRE_PDF(28)>mkdir extracted



D:\副研究員\Lip\20240903\ECG python\extracted\GP3_PRE_PDF(53)>mkdir extracted

一次把相同目錄內所有的 pdf 檔案都轉成 excel 放在” extracted” 中→

D:\副研究員\Lip\20240903\ECG

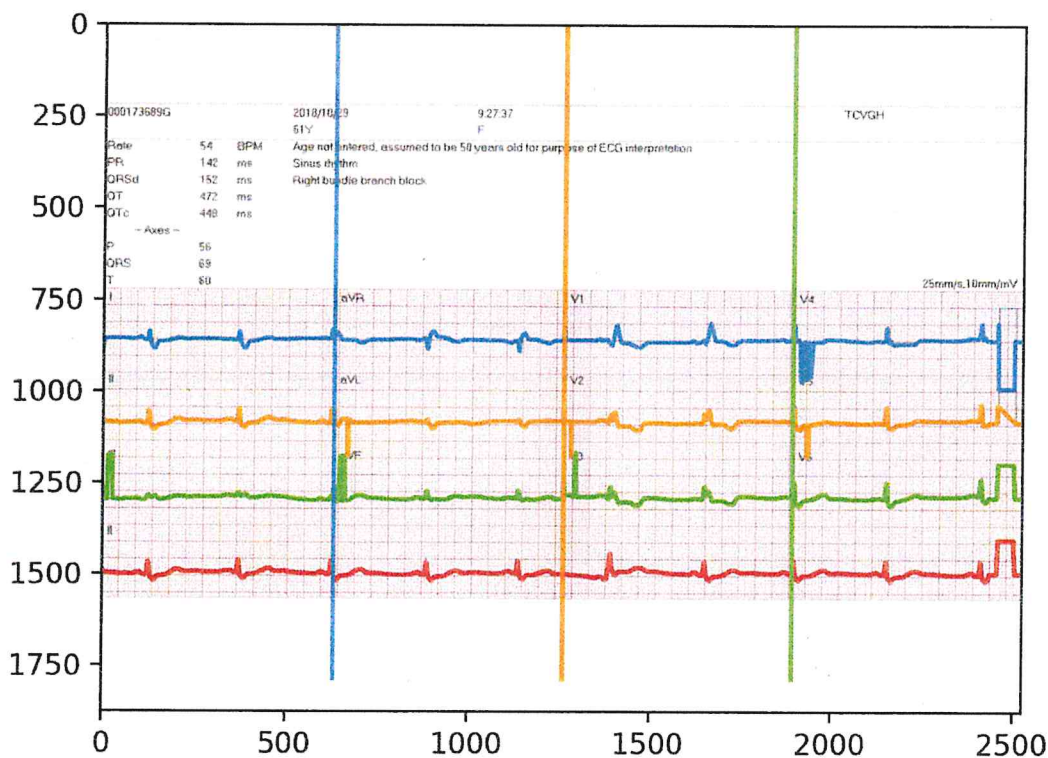
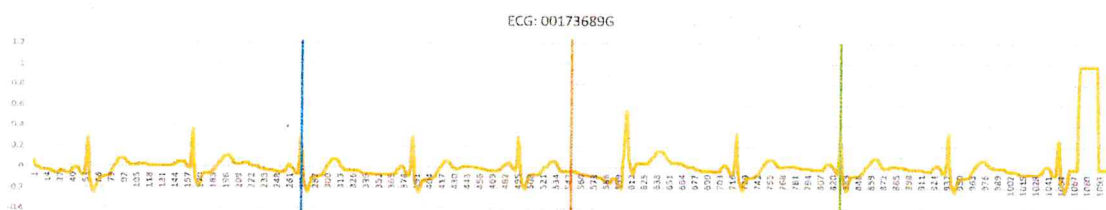
python\extracted\GP2_PRE_PDF(28)>python ..\raster_pdf_to_csv.py

python\extracted\GP3_PRE_PDF(53)>python ..\raster_pdf_to_csv.py

打開單一輸出的 Excel 隨每個 lead 時間切割點的數據位置，可以轉數字成 ECG 波型→

000173689G_2018-02-09 - Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Sample rate: 100 Hz																
2	Lead switches: 2.7489108910891087	5.502178217821782	783.825545544554455														
3	0.056679	0.028461	-0.00421	-0.00867	-0.00867	-0.01387	-0.01956	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045
4	0.056679	0.028461	-0.00421	-0.00867	-0.00867	-0.01387	-0.01956	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045
5	0.056679	0.028461	-0.00421	-0.00867	-0.00867	-0.01387	-0.01956	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045
6	0.056679	0.028461	-0.00421	-0.00867	-0.00867	-0.01387	-0.01956	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045	-0.03045



Python Code: raster_pdf_to_csv (python file)

```
import zlib
import sys
from locale import atof
import re
import os
import pymupdf
import numpy as np
from scipy.ndimage import *
from scipy.signal import find_peaks
import glob

from matplotlib import pyplot as plt

mmps, mmpmv = 25.0, 10.0
dpi=216
dpmm=dpi/25.4

clip_length=11.0 #seconds
target_hz=100

def reduce_lines(img):
    red=np.zeros(img.shape,dtype=np.uint8)
    for i in range(img.shape[1]):
        patch=False
        start=0
        for j in range(img.shape[0]):
            if patch and img[j,i]==0:
                red[int((j+start)/2),i]=255
                patch=False
            elif not patch and img[j,i]==255:
                start=j
                patch=True
        return red

def extract_waveforms(img):
    waveforms=[]
    for i in range(img.shape[1]):
```

```

for j in range(img.shape[0]):
    if img[j, i]==255:
        candidate_wave=[i, j]
        img[j, i]=0
        si=i
        for ii in range(i+1, img.shape[1]):
            if ii>si+60: break
            sj=j
            for jj in range(120):
                if sj+jj<img.shape[0] and img[sj+jj, ii]==255:
                    si=ii
                    candidate_wave.append(ii)
                    sj+=jj
                    candidate_wave.append(sj)
                    img[sj, ii]=0
                    break
                elif sj-jj>=0 and img[sj-jj, ii]==255:
                    si=ii
                    candidate_wave.append(ii)
                    sj-=jj
                    candidate_wave.append(sj)
                    img[sj, ii]=0
                    break

            if len(candidate_wave)>1000:
                waveforms.append(candidate_wave)

return waveforms

def find_lead_boundaries(img):
    img_b=np.where(img[:, :, 2]>200, 1, 0)*np.where(img[:, :, 0]<128, 1, 0)
    return find_peaks(np.sum(img_b, 0))[0]

def extract(filename, show=False):

    file=pymupdf.open(filename)
    file[0].set_cropbox(file[0].mediabox)
    pix=file[0].get_pixmap(dpi=dpi)
    img=np.frombuffer(pix.samples, dtype=np.uint8).reshape(pix.h, pix.w, pix.n)

```

```

img=np.ascontiguousarray(img)

lead_boundaries=find_lead_boundaries(img)

img_g=img[:, :, 0]
# plt.imshow(img_g, cmap='grey', interpolation='none')
# plt.show()

img_b=np.where(img_g<np.percentile(img_g,1.5), 255, 0)
# plt.imshow(img_b, cmap='grey', interpolation='none')
# plt.show()
img_b=median_filter(img_b, 3)
# plt.imshow(img_b, cmap='grey', interpolation='none')
# plt.show()
img_b=reduce_lines(img_b)
# plt.imshow(img_b, cmap='grey', interpolation='none')
# plt.show()

waveforms=extract_waveforms(img_b)

if show:
    plt.imshow(img)
    for i in range(4):
        print(len(waveforms[i]))
        plt.plot(waveforms[i][0::2], waveforms[i][1::2])

    for b in lead_boundaries:
        plt.plot([lead_boundaries[0], lead_boundaries[0]], [0, img.shape[0]])
        plt.plot([lead_boundaries[1], lead_boundaries[1]], [0, img.shape[0]])
        plt.plot([lead_boundaries[2], lead_boundaries[2]], [0, img.shape[0]])
    plt.show()

stretch=clip_length*dpmn*mmmps/(waveforms[0][-2]-waveforms[0][0])
lead_boundaries=(lead_boundaries-waveforms[0][0])*stretch/dpmn/mmmps
waves=np.empty((4, int(clip_length*target_hz)))
for i in range(4):
    waveforms[i][0::2]=[stretch*(x-waveforms[i][0])/dpmn/mmmps for x in
waveforms[i][0::2]]

```

```

        waveforms[i][1::2]=[-stretch*(x-np.mean(waveforms[i][1::2]))/dpm/mmpmv for x in
waveforms[i][1::2]]

```

```

waves[i]=np.interp(np.arange(0,0,clip_length,1.0/target_hz),waveforms[i][0::2],waveforms[i]
[1::2])

```

```

    if not show:
        csv = open('extracted/'+'.'.join(filename.split('.')[0:-1])+".csv","w")
        csv.write('Sample rate: '+str(target_hz)+' Hz\n')
        csv.write('Lead switches: '+str(lead_boundaries[0])+' '+str(lead_boundaries[1])+'
'+str(lead_boundaries[2])+'\n')
        for w in waves:
            for v in waves[i]:
                csv.write(str(v)+',')
            csv.write("\n")
        csv.close()

```

```

#main
if len(sys.argv)>1: extract(sys.argv[1],True)
else:
    for filename in glob.glob('*.*pdf'):
        print(filename)
        extract(filename)

```

附錄二

Title:

Long-Term Trajectories of Apolipoprotein A1 and Their Impact on Major Adverse Cardiovascular Events and Mortality

Background:

Apolipoprotein A1 (ApoA1) is a major component of high-density lipoprotein cholesterol and plays a critical role in reverse cholesterol transport. Dynamic changes in ApoA1 levels may be associated with major adverse cardiovascular events. This study aimed to evaluate the impact of ApoA1 trajectory changes over three early assessments.

Methods:

Participants in the Chin-Shan Community Cardiovascular Cohort with dyslipidemia and receiving three early ApoA1 assessments were enrolled. Group-based multivariate trajectory modeling was used to classify participants into distinct trajectory groups after multivariable adjustment. The follow-up duration was from April 1990 to August 2022, long-term outcomes of MACE and mortality were evaluated.

Results:

A total of 1,080 participants were included (mean age 66.92 ± 10.45 years, 43.2% males). Participants were classified into four ApoA1 trajectory groups: Trajectory 1 (T1: low-bell shaped), Trajectory 2 (T2: low-stable), Trajectory 3 (T3: high-bell shaped), and Trajectory 4 (T4: J-shaped). The cumulative incidence of MACE ranked as follows: T3 (6.5%) < T1 (8.0%) < T4 (8.8%) < T2 (11.2%). As compared with reference group of T3, T2 and T4 had significantly higher risks of MACE (T2: HR 2.32, 95%

CI 1.09–4.96; T4: HR 2.32, 95% CI 1.01–5.32). However, there were no significant differences in all-cause mortality among T1, T2, and T3 compared with T4.

Conclusion:

The trajectory of ApoA1 significantly influenced MACE risk during long-term follow-up, particularly in the low-stable and J-shaped trajectory groups. Dynamic monitoring of ApoA1 may serve as a valuable tool for early risk stratification in high-risk populations, facilitating more individualized interventions.

Keywords:

Apolipoprotein A1, cardiovascular outcomes, trajectory modeling, major adverse cardiovascular events, all-cause mortality.

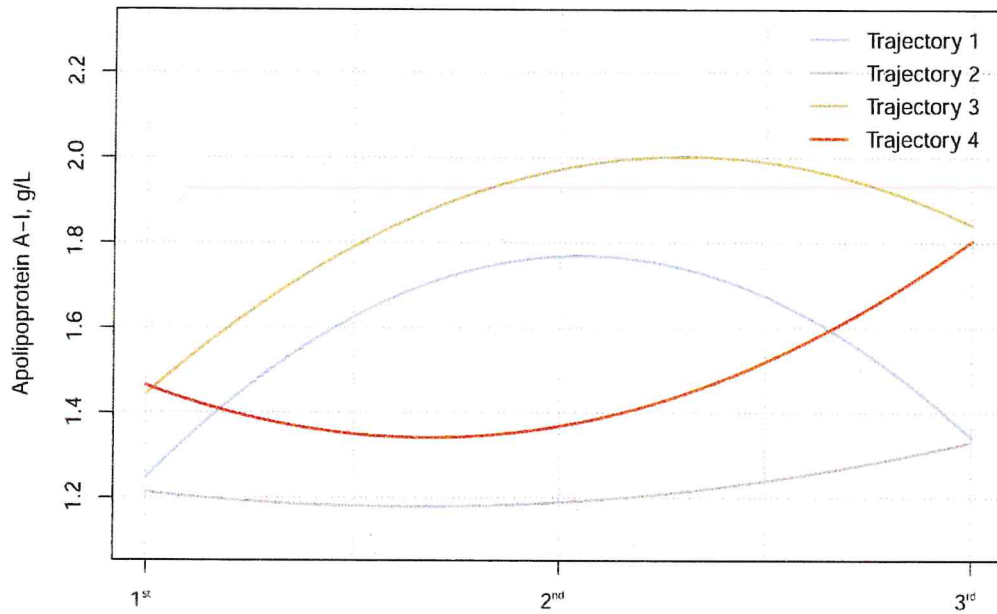


Figure. Classification of ApoA1 Trajectories Using Group-Based Multivariate Trajectory Modeling.

Group-based multivariate trajectory modeling classified participants into four distinct ApoA1 trajectory groups after multivariable adjustment: Trajectory 1 (T1: low-bell shaped), Trajectory 2 (T2: low-stable), Trajectory 3 (T3: high-bell shaped), and Trajectory 4 (T4: J-shaped).

Table. Associations of trajectories and different outcomes.

Outcomes	HR	95% CI	<i>P-value</i>
Composite outcomes			
Trajectory 1	1.12	0.82–1.52	0.475
Trajectory 2	1.22	0.91–1.64	0.189
Trajectory 3	<i>Reference</i>		
Trajectory 4	1.28	0.88–1.87	0.202
All-cause Death			
Trajectory 1	1.11	0.81–1.50	0.524
Trajectory 2	1.21	0.90–1.63	0.208
Trajectory 3	<i>Reference</i>		
Trajectory 4	1.27	0.87–1.85	0.225
Cardiovascular Death			
Trajectory 1	1.43	0.54–3.80	0.471
Trajectory 2	3.24	1.33–7.86	0.010
Trajectory 3	<i>Reference</i>		
Trajectory 4	3.61	1.32–9.88	0.013
Non-cardiovascular Death			
Trajectory 1	1.06	0.77–1.47	0.725
Trajectory 2	1.01	0.74–1.39	0.942
Trajectory 3	<i>Reference</i>		
Trajectory 4	1.02	0.67–1.56	0.925
MACE			
Trajectory 1	1.48	0.68–3.22	0.326
Trajectory 2	2.32	1.11–4.82	0.025
Trajectory 3	<i>Reference</i>		
Trajectory 4	2.52	1.07–5.91	0.034

MACE: major adverse cardiovascular events.

附錄三

This study utilized a dataset comprising 1,108 patients with medical records collected across three time periods: 1990–1991, 1992–1993, and 1997–2006. The trial spanned 390 months, from April 1990 to August 2022. We employed a deep learning model, Dynamic-DeepHit [1], to predict cardiovascular disease (CVD) mortality risk over time.

The model was developed using 14 biomarkers, including 3 binary features—sex, CAD history, and CVA history—and 11 continuous features, namely age, BMI, TG, CHOL, HDL, LDL, GLU, UA, WBC, ApoB, and ApoA1.

For model evaluation, the dataset was split into training and testing sets with a ratio of 8:2 (886:222). To investigate the relationship between features and causes of death, we randomly selected 60 patients, including 20 who died of CVD, 20 who died of other causes, and 20 who were censored for visualization. The risks for these patients were then derived and plotted using the model trained on the remaining patients.

The experiment was conducted in a Python 3.7 environment using TensorFlow 1.13. The model was trained for 50,000 epochs with a learning rate of $1e-4$. Long Short-Term Memory (LSTM) [2] networks formed the backbone of the model architecture, allowing for the capture of temporal dependencies within the patient data.

We predicted CVD mortality risk at six time points: 216, 252, 288, 324, 360, and 396 months from the start of the study. The model's output represented the estimated risk of CVD-related death for each patient at each evaluation time. Performance was assessed on the testing set using two metrics: the concordance index (C-index) and the Brier score. For patients with CVD as the cause of death, the model achieved a C-index of 0.9491 at 216 months and 0.9173 at 252 months.

Ref:

1. Lee, Changhee, Jinsung Yoon, and Mihaela Van Der Schaar. "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data." *IEEE Transactions on Biomedical Engineering* 67.1 (2019): 122-133.
2. Hochreiter, S. "Long Short-term Memory." *Neural Computation MIT-Press* (1997).

Survival Analysis based on Deep Learning

Environment establishment

```
conda create -n tensorflow1_13 python=3.7
conda activate tensorflow1_13
conda install numpy==1.16.5 pandas==1.0.1 scikit-learn==0.22.1
lifelines==0.24.9 termcolor==1.1.0
conda install -c sebp scikit-survival==0.12
conda install tensorflow=1.13
Not GPU tensorflow!!!!
```

The column name of input and output

Path =

/media/ruojun/X10Prol/Dynamic-DeepHit-master/data/CCCC_2024_data_CY_20240904.xlsx

Input columns = []

Dataset preprocessing

2437 patients have 0, 2nd, 5th records.

1108 patients don't have nan value.

Model: Dynamic Deephit

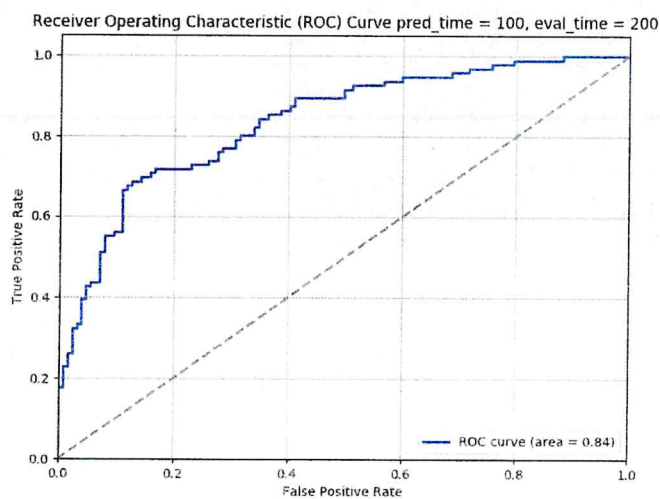
Results

Results of binary death label

- C-INDEX:				
	eval_time 70	eval_time 100	eval_time 200	eval_time 300
pred_time 70: event_1	0.832579	0.810093	0.862022	0.801705
pred_time 100: event_1	0.810093	0.854889	0.843552	0.802100
pred_time 150: event_1	0.869447	0.860506	0.812694	0.802468

- BRIER-SCORE:				
	eval_time 70	eval_time 100	eval_time 200	eval_time 300
pred_time 70: event_1	0.006044	0.030516	0.096764	0.192448
pred_time 100: event_1	0.030369	0.051578	0.125000	0.188642
pred_time 150: event_1	0.060934	0.075786	0.161509	0.448339
=====				

Use data from patients who are alive after 70 months to predict the prediction performance at 100th month



Results of new data

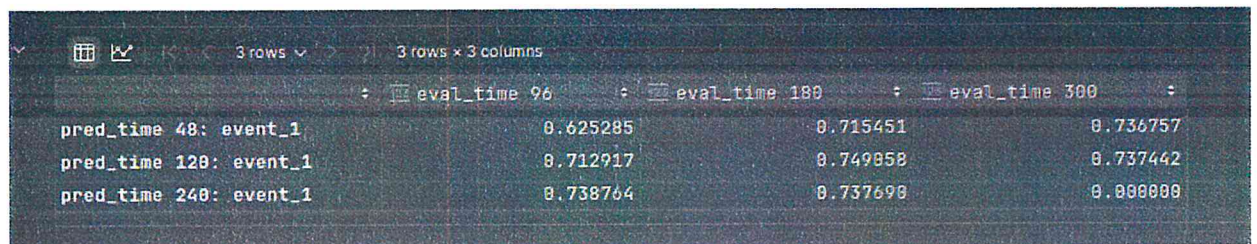
2437 patients have 3 records.

1108 patients have 3 full records.

Pred_time: 存活到當下的時間點

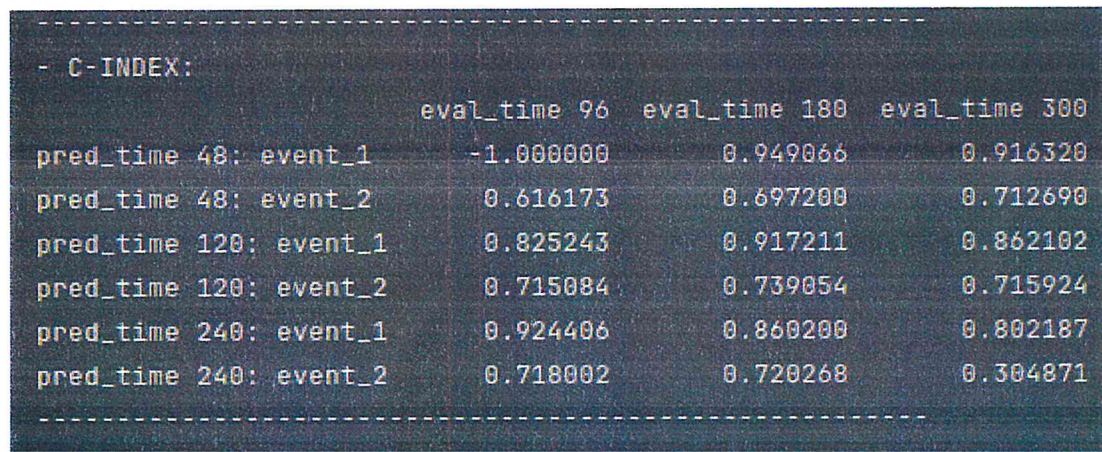
Eval_time: 已存活到當下的時間點去預測未來更遠的時間點

All-cause death (0/1)



	eval_time 96	eval_time 180	eval_time 300
pred_time 48: event_1	0.625285	0.715451	0.736757
pred_time 120: event_1	0.712917	0.749858	0.737442
pred_time 240: event_1	0.738764	0.737698	0.888888

Death causes (0: alive; 1: CVD, 2 : other death)



	eval_time 96	eval_time 180	eval_time 300
pred_time 48: event_1	-1.000000	0.949066	0.916320
pred_time 48: event_2	0.616173	0.697200	0.712690
pred_time 120: event_1	0.825243	0.917211	0.862102
pred_time 120: event_2	0.715084	0.739054	0.715924
pred_time 240: event_1	0.924406	0.860200	0.802187
pred_time 240: event_2	0.718002	0.720268	0.304871

0/1/2 death causes (6 time points)

Death causes (0: alive; 1: CVD, 2 : other death)

Pred_time: 存活到當下的時間點

Eval_time: 已存活到當下的時間點去預測未來更遠的時間點

時間單位：月

C-index						
	eval_time 216	eval_time 252	eval_time 288	eval_time 324	eval_time 360	eval_time 396
pred_time 60: event_1	0.949	0.917	0.915	0.850	0.851	0.860
pred_time 60: event_2	0.725	0.716	0.713	0.712	0.705	0.634
pred_time 120: event_1	0.924	0.851	0.853	0.857	0.780	0.780
pred_time 120: event_2	0.721	0.714	0.720	0.706	0.312	0.312
pred_time 204: event_1	0.858	0.852	0.768	0.768	0.768	0.768
pred_time 204: event_2	0.707	0.600	0.306	0.306	0.306	0.306
Score						
	eval_time 216	eval_time 252	eval_time 288	eval_time 324	eval_time 360	eval_time 396
pred_time 60: event_1	0.016	0.022	0.030	0.049	0.080	0.128
pred_time 60: event_2	0.082	0.107	0.135	0.180	0.219	0.275
pred_time 120: event_1	0.027	0.036	0.067	0.111	0.154	0.154
pred_time 120: event_2	0.126	0.155	0.206	0.248	0.305	0.305
pred_time 204: event_1	0.083	0.136	0.164	0.164	0.164	0.164
pred_time 204: event_2	0.218	0.270	0.298	0.298	0.298	0.298

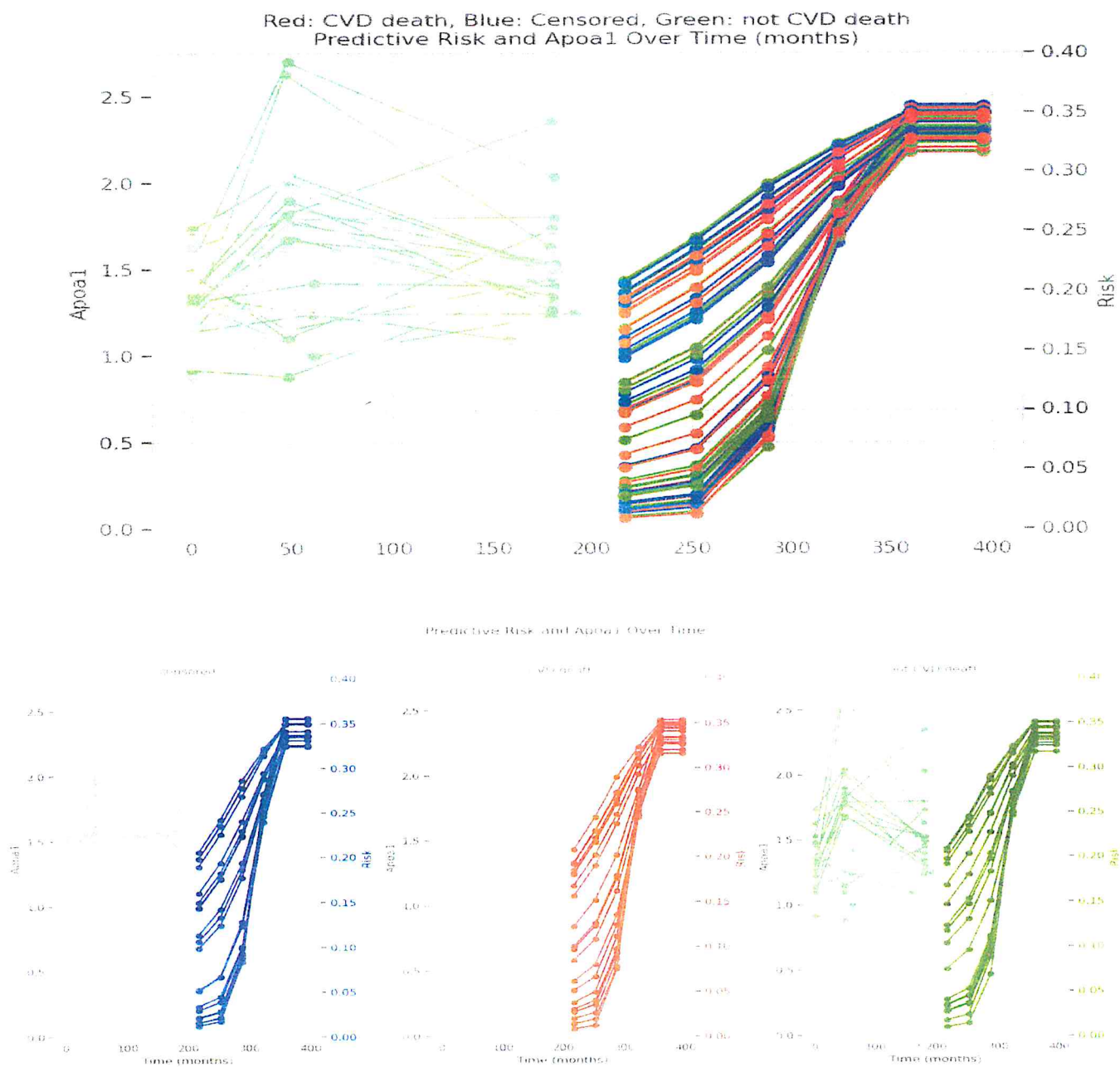


Figure: Trend of ApoA1 and predictive risks over time