

心肌梗塞預後分析系統

專題編號：109-CSIE-S002

執行期限：108 年第 1 學期至 109 年第 1 學期

指導教授：白敦文

專題參與人員： 106590012 彭柏誠、106590019 簡旭均
106590034 吳陽生、106590036 梁博瑜

一、摘要

心血管疾病在全球的主要死因排名位居第一，其中又以心肌梗塞最需注意。在本次的專題中，我們使用機器學習以及資料分析的相關技術，配合台灣蒐集超過 25 年以上的健保資料庫，完成一個心肌梗塞的預後系統。主要針對心肌梗塞患者確診後可能面臨的不同情況提供參考。

關鍵詞：心肌梗塞、預後分析、健保資料庫、機器學習。

二、緣由與目的

心血管疾病每年在全球造成超過 1790 萬人的死亡，佔全球總死亡人數的 31%，是全球主要死因排名的首位。根據估計，2030 年全球因心血管疾病死亡的人數將攀升到相當於整個台灣的人口數—2,300 萬人。另外根據衛服部的資料顯示：2014 年有近 2 萬國人因心臟疾病去世，平均每 27 分鐘就有一人死亡；從近 5 年來看，死亡率更是有 5.9% 的上升。由此可見，心血管疾病死亡率的上升已是全球明顯趨勢，成了不得不重視的問題。

在心血管疾病當中，又屬心肌梗塞最需注意。即使這十幾年來醫學的精進，降低了心肌梗塞的死亡率，但隨著台灣社會工業化、國人飲食西化、空氣汙染及現代緊張的生活節奏，國內心肌梗塞的發生率有逐年升高並年輕化的趨勢。

在這次的專題中，我們利用機器學習以及資料分析的技術，完成一個心肌梗塞預後分析系統。相較於以往的預後分析大多只依靠病患的體重、年齡、性別等數據

進行簡單的統計分析[1]，最後由醫生來進行判斷，這樣的結果容易因為醫生的經驗不一而有所不同，難以有統一的標準。我們希望從人工智慧的角度切入，利用台灣收集超過 25 年以上的健保資料庫，提供更具有參考價值的預後分析。

三、技術方法

(一) 隨機森林(Random forest)

隨機森林模型透過決策樹預測器的集合進行預測，以獨立隨機選擇的特徵以及相同的參數設定訓練集合中的所有決策樹。此種方法已經被證明在特徵較多且在樣本數量不至過少時，可以獲得良好的表現，且不易受到冗餘特徵的影響[2]。

(二) XGBoost

XGBoost 是一個開源的庫，實現了梯度提升樹的算法。梯度提升樹是一種 boost tree 的優化方案，透過梯度下降的概念，使每一次 boost 的過程中損失函數能夠盡可能的變小，進而提升預測模型的準確度。在 2015 的 Kaggle 機器學習競賽中，29 個項目裡面有 17 個項目的優勝組使用了 XGBoost。

(三) Logistic regression

是一種基於勝算比的自然對數所建構的預測模型，可以用來解決線性回歸中的兩個問題：第一，極端直並不會符合線性關係；第二，錯誤不會均勻的分佈在資料群中。建構過程中，採用 LBFGS 演算法優化，將最大迭代次數設為 100，同時使用 L2 正規化減少模型過擬合的可能性。

(四) Apriori

用來在關聯規則學習過程中提取頻繁模式。透過事先設定好的支持度閾值，

利用 downward closure 的特性，一層一層的產生候選組合來找出所有的結果。

(五) Flask

使用 Python 編寫的輕量級 Web 應用框架。負責提取資料運算，並與前端進行連接。

(六) MPAndroid

第三方繪製圖表套件，用來視覺化呈現後端的運算結果。

(七) OKHttp

第三方網路連線套件，向 server 端發送請求並獲取用戶的個人資料。

四、系統簡述

(一) 短期內死亡風險預測模組

在研究過程中我們發現：心肌梗塞的確診病患中，大量的患者在確診一周內即過世，佔了死亡總人數的 15%。對於心肌梗塞病患而言，確診後的七天是最為關鍵的。為此，我們訓練了一個短期內死亡的風險預測模型，針對那些高風險的族群，期望可以提早發現並給予適當的治療與照護，想辦法幫助患者度過這段時間。

(二) 併發心臟衰竭風險預估模組

根據研究，加拿大艾柏塔省(Alberta)在 1994 年到 2000 年間的病患中，有高達 76% 的心肌梗塞病患會同時罹患心臟衰竭 [3]。而從台灣的健保資料庫中也能發現同樣的情況：有約 5 成的病患在罹患心肌梗塞後併發心臟衰竭。針對心肌梗塞確診後併發心臟衰竭的可能性，我們訓練了一個模型進行評估。

(三) 疾病關聯規則學習模組

對於心肌梗塞患者而言，除了要注意死亡風險最高前七天，在之後的一年內仍然是非常關鍵的一段時間。根據我們的資料顯示：心肌梗塞患者在確診一年內死亡的人數高達 961 人，其中有 335 人為 7 日內死亡的患者，扣除這些急性心肌梗塞案例，仍然有 626 人在第一年死亡；相較於確診後第二年到第六年的平均死亡人數 (135.2) 多了約 4.63 倍。因此我們在系統中同時引入了疾病關聯規則學習的模組，挖

掘出患者在確診後一年內可能患有的疾病規則，希望可以協助患者提早進行預防與治療，降低死亡風險。

(四) 手機端應用程式

為了方便患者使用，除了機器學習的相關模型，我們透過手機端的應用程式來完成資料的視覺化呈現。使用者登入帳號並授權健康存摺資料的使用後，所有患者的資料將會進行統計分析與進入機器學習模型進行評估，經過整理後再以視覺化的方式呈現在前端。

五、結論

除了基本的統計分析外，整個系統從患者確診後可能遇到的問題，分為三個時間段，透過不同的機器學習技術提供患者相關的參考資訊。對於每個子模組的內容，我們也深入地做了相關的文獻調查與研究。目前關於併發心臟衰竭風險預估模組的成果，已經整理成論文於 DLT2020 完成發表；而剩下兩個模組的研究成果，目前也正積極準備撰寫論文並發表。

參考文獻

- [1] Thomas, K.L., Velazquez, E.J, "Therapies to prevent heart failure post-myocardial infarction." Curr Heart Fail Rep 2, 174–182 (2005).
- [2] Tin Kam Ho, "The random subspace method for constructing decision forests", IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 20, Issue: 8, Aug 1998).
- [3] Torabi A, Rigby AS, Cleland J, "Declining in-hospital mortality and increasing heart failure incidence in elderly patients with first myocardial infarction." J Am Coll Cardiol 2009;55:79–81.