

建構以大數據為基礎的併發症預警模型-以高血脂併發症為例

王乃弘、李曉慧、張肅婷、林政杰、林俊榮、王思怡

摘要

三高疾病(高血壓、高血脂、高血糖)所引發的併發症經常是不可逆的慢性疾病或是會對生命造成危害的心血管疾病，而我國的十大死因主要即是以慢性疾病為主，依序為惡性腫瘤、心臟疾病、肺炎、腦血管疾病等，而最容易引起罹患這些慢性疾病原因就是三高疾病。因此如果能對有較高併發症發生風險的三高病患及早預警，並得以實施早期衛教，應有助於降低併發症發生的機率。

與大數據有關的資料科學應用近年來也是頗受重視的熱門議題，而人工智慧的發展更是近期政府和產業積極推動和投入的領域和範疇。本研究即希望能藉由資訊相關技術資料探勘、分類技術與決策樹，透過大數據的分析，建置一個高血壓併發症預警系統，除了可以協助醫院及民眾能即早發現和改善外，後續能降低發病後所使用的醫療資源。

關鍵詞：高血脂、併發症、決策樹、大數據分析、預警系統

王乃弘，光田醫療法人光田綜合醫院院長室院長。E-mail: admin@ktgh.com.tw

李曉慧，台南應用科技大學連鎖加盟經營管理學士學位學程助理教授，E-mail: xiasohui@gmail.com

張肅婷，光田醫療法人光田綜合醫院院長室副院長。E-mail: finance@ktgh.com.tw

林政杰，光田醫療法人光田綜合醫院資訊部主任。E-mail: it.director@ktgh.com.tw

林俊榮*(通訊作者)，中臺科技大學人工智慧健康管理系副教授兼系主任。E-mail:

raymond@ctust.edu.tw

王思怡，國立彰化師範大學行銷與流通管理碩士班研究生。E-mail: doris19970919@gmail.com

Using Big Data Technology to Establish Complication Early Warning Model - Take Hyperlipidemia Complications as an Example

Nai-Phon Wang & Hsiao-Hui Li & Sue-Ting Chang

Cheng-Chieh Lin & Chun-Jung Lin & Wang, Sih-Yi

Abstract

Complications caused by the three high diseases (hypertension, hyperlipidemia, and hyperglycemia) are often irreversible chronic diseases or cardiovascular diseases that are life-threatening. The top ten causes of death in my country are mainly chronic diseases. Mainly, malignant tumors, heart disease, pneumonia, cerebrovascular disease, etc. are in order, and the most likely cause of these chronic diseases is the three high diseases. Therefore, if early warning can be given to patients with high risk of complications and early health education can be implemented, it should help reduce the probability of complications.

The application of data science related to big data has also been a hot topic that has received considerable attention in recent years, and the development of artificial intelligence is a field and category that the government and industry have actively promoted and invested recently. This research hopes to build an early warning system for hypertension complications through the analysis of information-related technical data, classification techniques and decision trees, and through the analysis of big data. In addition to helping hospitals and the public to detect and improve immediately, follow-up Can reduce the medical resources used after the onset of disease.

王乃弘、李曉慧、張肅婷、林政杰、林俊榮、王思怡 建構以大數據為基礎的併發症預警模型-以高血脂併發症為例

Keywords: hyperlipidemia, complications, decision tree, big data analysis, early warning system

Nai-Phon Wang, Superintendent, Kuang Tien General Hospital, Email: admin@ktgh.com.tw

Hsiao-Hui Li, Assistant Professor, Bachelor's Degree Program in Chain Store Management, Tainan University of Technology, Email: xiasohui@gmail.com

Sue-Ting Chang, Vice President, Kuang Tien General Hospital, Email: finance@ktgh.com.tw

Cheng-Chieh Lin, Director of Information Department, Kuang Tien General Hospital, Email: it.director@ktgh.com.tw

Chun-Jung Lin* (Corresponding Author), Associate Professor, Chairman, Dept. of AI and Health Management, Central-Taiwan University of Science and Technology, Email: raymond@ctust.edu.tw

Wang, Sih-Yi , Student, Dept. of Marketing and Logistics Management, National Changhua University of Education, E-mail: doris19970919@gmail.com

壹、前言

隨著國人的飲食精緻化與西化、3C 的普及應用，國人的運動量與活動量普遍降低，這樣的生活飲食習慣的改變，使得國人提早且提高了罹患慢性疾病的機率。現今許多疾病對健康所造成的威脅，經常是令人措手不及的。成大醫院心臟內科謝棟漢醫師表示，依據衛生福利部公佈之國人 10 大死因統計(衛生福利部，2020)，依死亡率排序，108 年十大死因依序為：(1)惡性腫瘤(癌症)、(2)心臟疾病、(3)肺炎、(4)腦血管疾病、(5)糖尿病、(6)事故傷害、(7)慢性下呼吸道疾病、(8)高血壓性疾病、(9)腎炎腎病症候群及腎病變、(10)慢性肝病及肝硬化，排名順位與 107 年相同。而一般民眾普遍認知與「三高」有關係的就有(2)心臟疾病、(4)腦血管疾病、(5)糖尿病與(8)高血壓性疾病等四種。

預防勝於治療，這是人們耳熟能詳的一句話，疾病發生後除了增加醫療費用的支出外，不論疾病治療後的改善情況如何，整體的身心健康狀態或多或少都會被影響，而對於許多不可逆的狀況(例如：截肢)，伴隨的疾病的發生與治療，更會影響到人們的生活品質。在三高疾病也早已經不再是老年人的專利的情況下，許多疾病在發生前，我們的身體會發出疾病的「前兆」，若能將這些徵候與資訊進行適當的分析與判斷，應更有機會讓患者把握黃金治療時期，除了可以降低病發後所需花費的時間與費用外，更也有機會讓身心健康狀況維持在較好的狀態。

人工智慧在西元 1997 年由 IBM 公司所設計的深藍電腦(Deep Blue)贏了人類的西洋棋王(Garry Kasparov)開始就已經開始有相當多的討論，到了 2017 年由 Google DeepMind 設計的 AlphaGo 在中國烏鎮圍棋峰會擊敗以九段棋王柯潔為首的對奕比賽後，更是將人工智慧的研究與應用帶到另一個高峰，我國政府和產業更也是積極推動和投入人工智慧的領域和範疇。近年來以大數據相關的發展與應用是頗受重視的熱門議題，而大數據的本質是資料分析，更也是人工智慧的重要基礎。以機器學習來看，機器學習即是從過去的資料分析，去預測或反應未來可能出現或發生的狀況或結果。

本研究以概念驗證(Proof of Concept)的方式，透過大數據的分析建置一個高血脂併發症預警系統，除了可以協助醫院及民眾能即早發現和改善外，也能降低併發疾病發生後所使用的醫療資源。本研究利用醫療院所的醫院資訊系統，將個資進行適當的刪除與遮蔽後，利用文字探勘萃取出病患在確診前出現頻率較高的主他訴症狀、檢驗結果以及處置等，將這些資料建置成一個疾病預測模型系統。醫療院所可以將民眾相

關的就診紀錄帶入疾病預測模型系統中，以計算出病患發生高血脂併發症的可能機率，讓醫療院所可以協助民眾提早發現並可提早進行衛教與治療，以延緩發病的時間。

貳、文獻探討

本研究建置以大數據分析與資料探勘為基礎的高血脂併發症預警系統，以下將分別針對高血脂及相關之資訊技術進行文獻探討與說明。

一、高血脂與其併發症

高血脂症的定義：血液中所含的脂肪，主要包括：膽固醇、三酸甘油酯、磷脂質及游離脂肪酸等，一般總血膽固醇值或血三酸甘油酯值超過 200mg/dl 者即可診斷為高血脂症(彰化基督教醫院血管醫學防治中心資訊網，2010)。表 1 為成人膽固醇及三酸甘油酯的濃度參考值，以下數據任一個異常或是合併多整異常皆可稱高血脂或血脂異常。

表 1、成人膽固醇及三酸甘油酯參考濃度值(彰化基督教醫院血管醫學防治中心資訊網，2010)

禁食12小時	理想濃度	邊際危險濃度	高危險濃度
總膽固醇(非禁食)	<200mg/dL	200~239mg/dL	>=240mg/dL
低密度脂蛋白膽固醇	<100mg/dL	130~159mg/dL	>=160mg/dL
高密度脂蛋白膽固醇	>=60mg/dL		<40mg/dL
三酸甘油酯	<150mg/dL	200~400mg/dL	>400mg/dL

表 2 為高血脂併發症主要受害的器官及因素，血脂與心血管疾病有著相互影響的關係，血脂過多，產生斑塊，阻塞血管(問上醫，2017)。臨床上，約有一半以上的心肌梗塞患者，病發前會出現持續性胸痛、盜汗、噁心和嘔吐等典型的病徵(吳珮均，2017)，但大多數人會沒有自覺，以為是其他的病兆，嚴重者會昏厥、低血壓、胸痛、無心跳、甚至死亡。

表 2、高血脂併發症

主要受害器官	疾病	因素
心臟	冠心病	當血脂過多，易造成血稠，在血管壁上容易沉積，逐漸成小斑塊，斑塊若增多，阻塞血管，使血流變慢。
腦	腦中風	
眼睛	失明	
腎臟	腎動脈硬化 腎功能衰竭	
下肢	肢體壞死、潰爛	
胰臟	糖尿病	胰島素供應不足，容易造成血脂升高。
肝臟	脂肪肝	肝臟功能不佳者會影響脂蛋白作用，伴隨血脂代謝異常。

根據醫學研究指出，溫度對血管造成的變化，稱為「U型曲線現象」，當溫度過冷或過熱，都會影響血管收縮或擴張（魯皓平，2017）。

二、大數據

近年來大數據的運用愈來愈廣泛，常見的大數據應用領域，如在商業上：用來做決策分析、市場銷售之預測、設計新的銷售方式等，在醫學上：利用電子病歷、就診相關資料等做分析和預測等，在科學上：利用大量的地理資訊和環境相關資料來做一些預警，如防洪、地震等等(蔡毓娟，2015)。大數據的運用解決不少在海量資料及數據上的任何問題。本研究以疾病預測為主軸，收集發病患者確診前的病歷加以分析，以獲得發生疾病前會有的症狀、檢查報告、檢驗結果、使用過的藥物以及相關的處置。

使用大數據進行預測分析的案例中，Google 流感趨勢預測 (Google Flu Trends, GFT)是相當著名的案例。Google 在西元 2008 年時，以大數據分析來預測流感可能的發生地圖，其預測結果據稱是與美國疾病預防管制中心十分接近。它是利用像是使用「咳嗽」、「發燒」等流感相關的字詞進行搜尋的關鍵字數據進行預測分析，並快速的預測流感可能從何地開始出現及傳播的範圍。Google 宣稱其系統成功預測了 H1N1 在全美的傳播狀況，但之後發現 GFT 的預測即不再準確，許多學者與機構也對此現象提出可能的謬誤或迷思，葉淑婷(2013)在其研究中指出，對於使用 Google 進行流感相關議題的搜尋者可以區分為兩類，一類是對流感議題有興趣的，另一類則是感

冒或流感病患者，而後者的搜尋所產生的數據才是對預測有用的，因利用 Google 搜尋關鍵字進行疾病預測還是會有失真的情況。

三、決策樹

所謂的決策樹是指一個樹狀結構：樹的中間節點(non-leaf nodes)代表測試的條件，樹的分支(branches)代表條件測試的結果，而樹的葉節點(leaf nodes)則代表分類後所得到的分類標記，也就是表示分類的結果。自 1960 年代開始，已有許多學者使用樹狀結構來進行資料分析，包括 ID3、C4.5 等，ID3、C4.5 只適用於類別變數，如有連續性屬性必須先進行資料轉換。決策樹是應用廣泛的分類與預測工具。決策樹是以樹狀圖為基礎的方法，具有讓使用者容易理解的優點。一般來說，決策樹的準確性是依據資料來源的多寡，如果決策樹是透過了龐大的資料集所建構出來，則其預測的結果通常是符合期望的(尹相志，2009)。

決策樹在醫療產業上也有許多研究，以類神經網路、支援向量機、決策樹等技術建立心血管疾病術後基礎的預測模型，接著再以組合的方式建立心血管預測模型，其結果將用來建立決策支援系統之用(蔡詩怡，2011)。建立慢性腎臟病各個分期的預測模型，讓醫生、病患得知腎臟病期數，如此便能降低台灣洗腎的比例，此研究是利用決策樹、類神經模型進行預測(黃原博，2012)。

本研究以決策樹來進行資料分類，先將資料依類別屬性進行分類，以樹狀圖的方式找出疾病的特徵，透過決策樹分類及特徵顯示，進行分析，將此結果來增加預測的準確率。

參、研究方法

本研究以中部某區域醫院之醫院資訊系統資料庫做為研究的資料來源，藉由大數據分析技術，建置以高血脂病人併發症為主的預警系統。本研究之研究流程如圖 1 所示，說明如下：

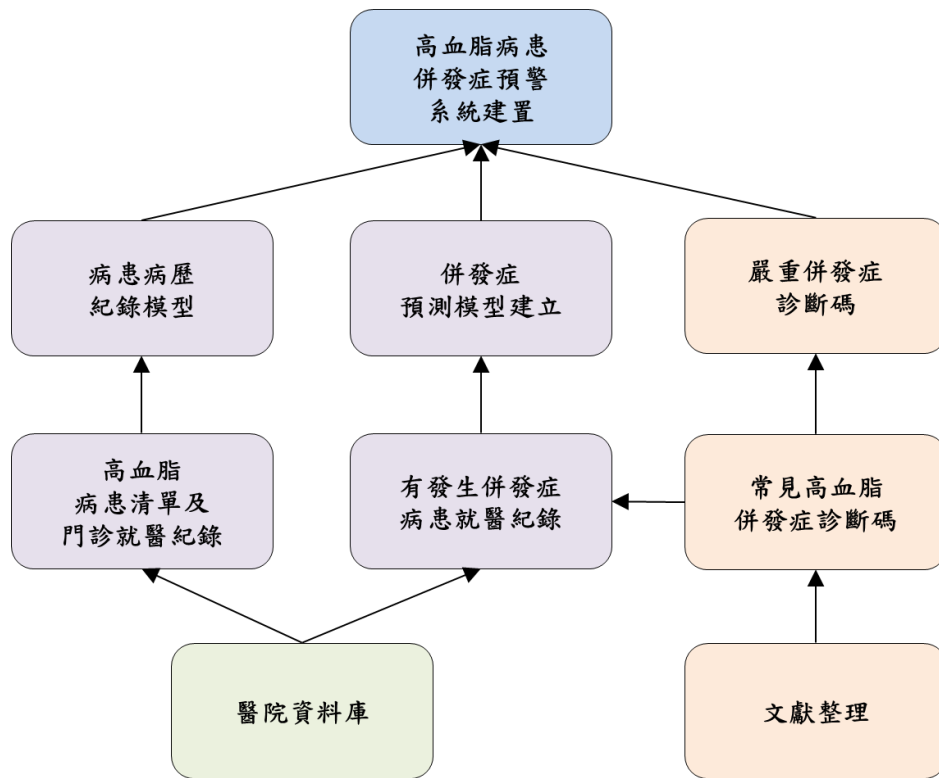


圖 1 研究流程圖

一、 併發症預測模型建立

1. 從文獻整理來發現常見的高血脂併發症，並建立常見併發症診斷碼知識庫。
2. 參考例如像是十大死因、術後存活率等資料，從常見併發症診斷碼中，挑選出嚴重併發症之診斷碼並建置嚴重併發症知識庫。
3. 依照常見併發症診斷碼知識庫，從醫院資料庫篩選出過去有罹患併發症的病患清單及其就醫紀錄。
4. 應用決策樹及分類技術，從併發症就醫紀錄的資料建立併發症發病預測模型。

二、 數據分析架構

圖 2 為數據分析流程圖，其運作模式為：

1. 將醫院實際門診資料從資料庫抓取病患就醫紀錄(主述、他述)、病歷資料(檢驗項目數據、用藥紀錄)有確診高血脂的病患抓取出來。

2. 利用資料探勘找出併發症的特徵、診斷碼，將以上資料建置成訓練數據，放入 Weka 訓練準確度。
3. 並以文獻探討及與醫師討論的疾病結果作為預測輔助，增加準確率及可信度。
4. 再將以上兩者數據經由權重比例進行計算，作出訓練模型。
5. 在預測時，將醫院系統中新病人的病歷資料、檢驗項目放入系統內，與已罹患高血脂的資料系統進行比對，並顯示罹患高血脂機率。
6. 若確定罹患高血脂，將檢驗數據與併發症的檢驗數據作為比對，在與醫師討論、文獻探討的結果交叉比對，並計算出高血脂併發症的預測機率。

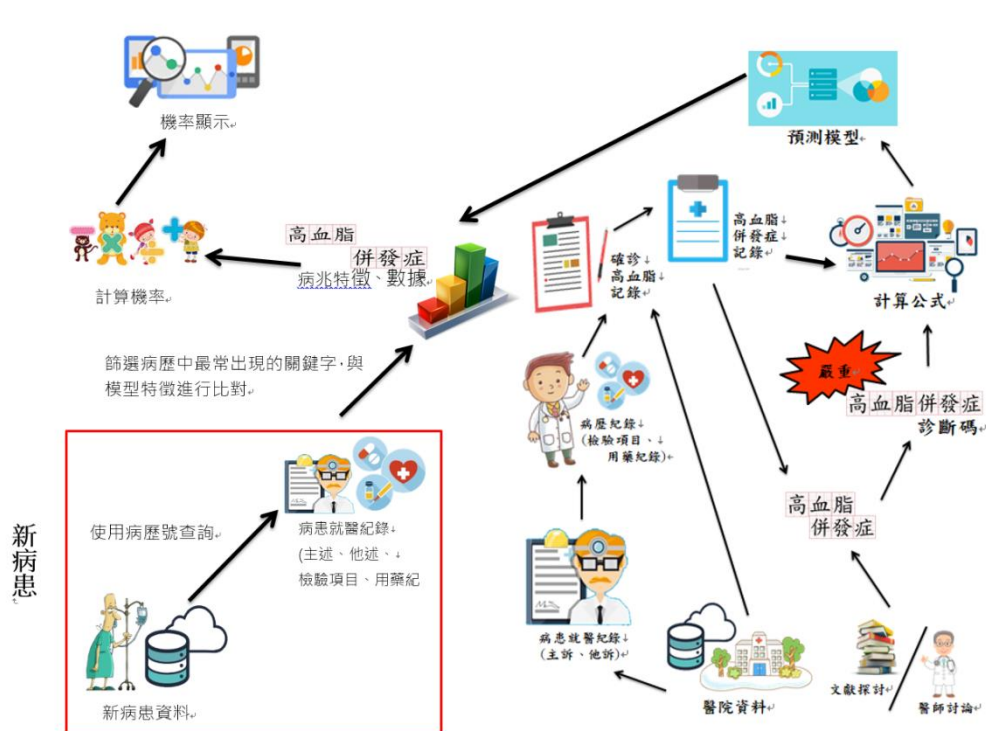


圖 2 數據分析流程圖

三、 模型製作

模型製作資料總筆數為 805,168 筆，透過資料庫將資料取出後進行資料正規化，將欄位分成病歷號、科別、主述、診斷碼、檢驗項目、檢驗數據，並將這些資料分成三分之二的訓練數據以及三分之一的測試數據。我們將訓練數據放入 Weka 進行模型

訓練，再將訓練結果製成決策樹以進行分析，且利用文字雲從主他述找出最常出現的特徵。我們透過 Weka J48 分類器製作出模型，透過每次的模型訓練、修改，讓模型的準確率越來越高，亦可提升本研究預測準確度。

四、 決策樹

我們將以上訓練資料進行分類，分別為總膽固醇、高密度脂蛋白、低密度脂蛋白、三酸甘油脂。將資料細分後，利用這些分類的資料進行決策樹分析，透過決策樹分類，能顯示出數據的區間，透過這些數據區間能夠了解容易罹患哪些疾病。

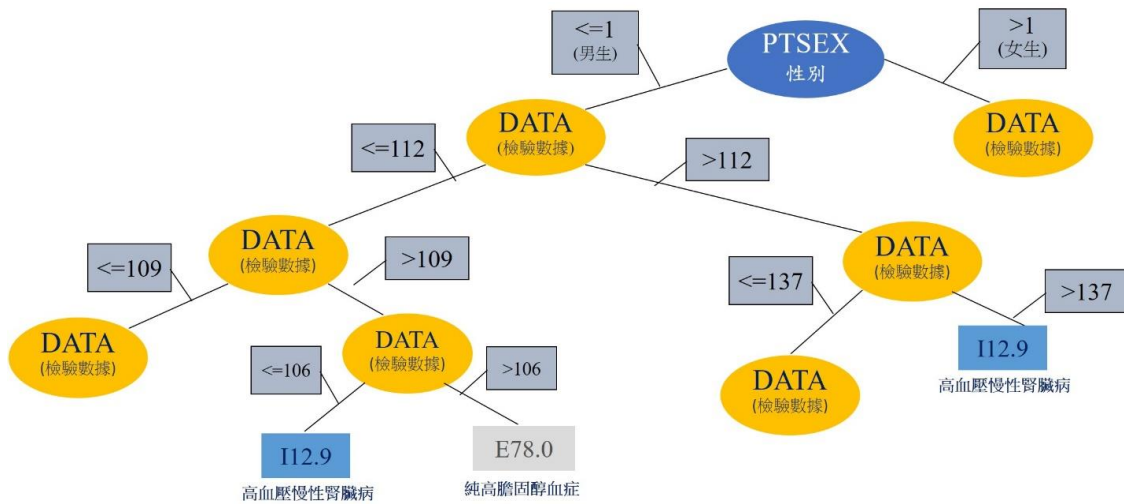


圖 3、總膽固醇決策樹部分畫面

圖 3 為總膽固醇決策樹分析說明。以性別(男生為 ≤ 1 ，女生為 > 1)為第一層分類，而第二層分類為總膽固醇檢驗數據，男生部分先以 112 作為分界值，如大於 112 時，又進行第三層的分類，以 137 作為分界值，而數據大於 137 時，則容易罹患診斷碼 I12.9(高血壓慢性腎臟病)；如數據小於等於 112 時，又進行第三層的分類，以 109 作為分界值，當大於 109 時，又進行第四層的分類，以 106 作為分界值，而數據大於 106 時亦容易罹患診斷碼 E78.0(純高膽固醇血症)，而小於等於 106 時容易罹患 I12.9(高血壓慢性腎臟病)。

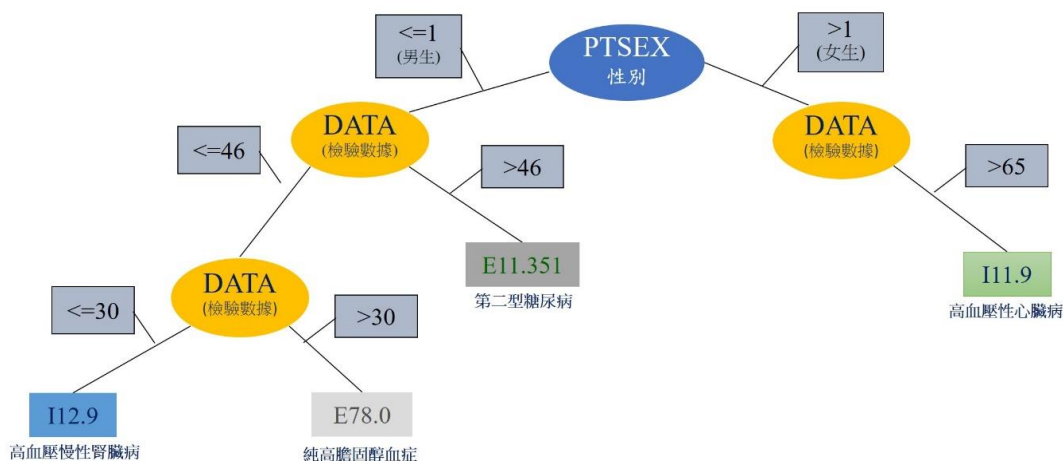


圖 4、高密度脂蛋白決策樹部分畫面

圖 4 為高密度脂蛋白決策樹分析說明。以性別(男生為 ≤ 1 ，女生為 > 1)為第一層分類，而第二層分類為高密度脂蛋白檢驗數據，男生部分先以 46 作為分界值，如大於 46 時，則容易罹患診斷碼 E11.351(第二型糖尿病)；如小於等於 46 時，又進行第三層的分類，以 30 作為分界值，而數據大於 30 時，則容易罹患診斷碼 E78.0(純高膽固醇血症)，而小於等於 30 時亦容易罹患診斷碼 I12.9(高血壓慢性腎臟病)；女生部分以 65 作為分界值，如數據大於 65 時亦罹患診斷碼 I11.9(高血壓性心臟病)。

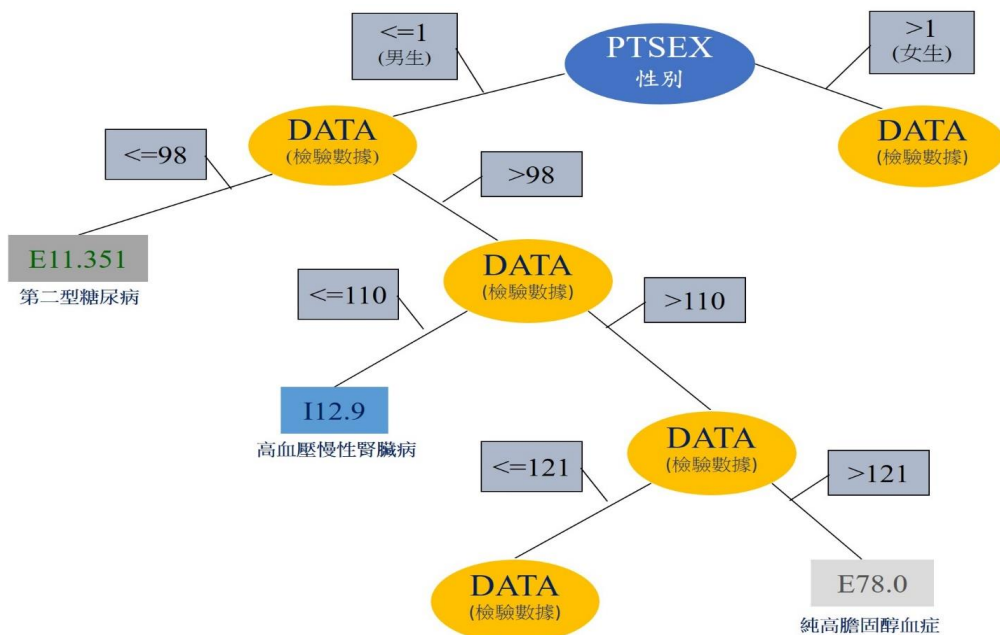


圖 5、低密度脂蛋白決策樹部分畫面

圖 5 為低密度脂蛋白決策樹分析說明。以性別(男生為 ≤ 1 ，女生為 > 1)為第一層分類，而第二層分類為低密度脂蛋白檢驗數據，男生部分先以 98 作為分界值，如小於等於 98 時，則容易罹患診斷碼 E11.351(第二型糖尿病)；如大於 98 時，又進行第三層

的分類，以 110 作為分界值，如小於等於 110 時，則容易罹患診斷碼 I12.9(高血壓慢性腎臟病)；如大於 110 時，又進行第四層的分類，以 121 作為分界值，而數據大於 121 時，則容易罹患診斷碼 E78.0(純高膽固醇血症)。

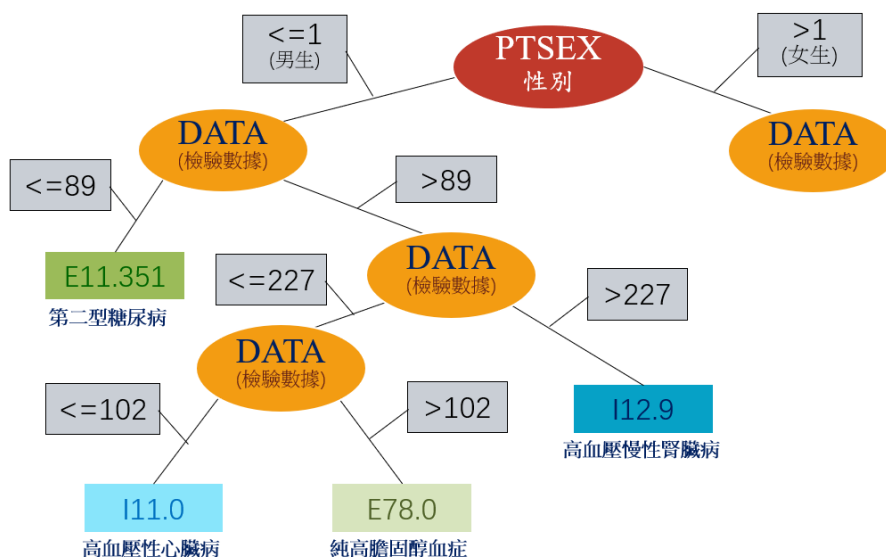


圖 6、三酸甘油脂決策樹部分畫面

圖 6 為三酸甘油脂決策樹分析說明。以性別(男生為 ≤ 1 ，女生為 > 1)做為第一層分類，而第二層分類為三酸甘油脂檢驗數據，男生部分先以 89 作為分界值，如小於等於 89 時，則容易罹患診斷碼 E11.351(第二型糖尿病)，而數據大於 89 時，又進行第三層的分類，以 227 作為分界值，當大於 227 時，亦容易罹患診斷碼 I12.9(高血壓慢性腎臟病)，而小於等於 227 時，進行第四次分類，以 102 最為數據分界值，小於等於 102 時容易罹患診斷碼 I11.0(高血壓性心臟病)，而大於 102 時容易罹患診斷碼 E78.0(純高膽固醇血症)。

五、 文字雲

透過文字雲將病患的主訴進行分析整理，並找出其文字詞彙出現的頻率值，文字詞彙經由排序就可得知出哪些檢驗出現最多次，我們也可以從詞彙清單刪除不必要的詞語，讓文字雲的分析更準確。文字雲也可以透過圖形方式顯示，透過視覺化的效果表示既簡單又明瞭，如圖 7 所示。

王乃弘、李曉慧、張肅婷、林政杰、林俊榮、王思怡 建構以大數據為基礎的併發症預警模型-以高血
脂併發症為例



圖 7、主述文字雲

肆、研究結果

本研究的研究結果，透過年齡的輸入、性別、發生的身體狀況以及身體數據，就能快速的了解有多少機率的機會會罹患併發症，且系統提供了衛教資訊，希望使用者能夠降低罹患併發症的風險。

圖 8 為系統執行畫面結果，年齡輸入 50，性別為女生，症狀有胸悶、肢體麻木等，總膽固醇值為 150，三酸甘油酯值為 130，高密度脂蛋白值為 80，低密度脂蛋白值為 110，按下預測，結果為右邊顯示，預測機率為 36%，屬於中度危險群，而下方有建議提供使用者做為參考，提醒使用者注意自身狀況，並建議使用者如何預防疾病的發生。

高血壓預測系統

年齡：

性別：
 男生 女生

症狀：
 頭暈 胸悶 健忘
 心悸 肢體麻木
 疲倦 失眠

總膽固醇 (CHOL)：

三酸甘油脂 (TG)：

高密度脂蛋白 (HDL)：

低密度脂蛋白 (LDL)：

併發症預測機率	分類
< 2.5 %	輕度危險
2.6 ~ 4.5 %	中度危險
> 4.6 %	重度危險

結果：

您輸入的資料如下：
年齡：50歲
性別：女
平時出現的症狀：
胸悶 肢體麻木
總膽固醇 (CHOL)：150 mg/dl
三酸甘油脂 (TG)：130 mg/dl
高密度脂蛋白 (HDL)：80 mg/dl
低密度脂蛋白 (LDL)：110 mg/dl
發生併發症預測機率：36%

您未來可能發生的併發症如下：
糖尿病

建議：
一、減少食用高熱量的食物，建議多攝取低膽固醇食品，如：五穀類、蔬菜水果。
二、若您為肥胖者，則必須減肥。
三、若您經常飲酒，請避免飲用過度。
四、若您為抽菸者，請減少抽菸量，若能戒菸極佳。
五、請養成運動習慣。



圖 8、系統執行畫面

伍、討論與建議

本研究將以大數據分析為基礎，透過人工智慧的相關技術建置高血壓併發症預警系統。本系統建置完成後，有以下的成效：

1. 人力方面：藉由此系統協助醫護人員在照護病患時，能獲得更多資訊，可以使病患獲得更積極的疾病發生預防和照護。
2. 系統方面：藉由該系統的建置完成，未來可以有效的應用到其他疾病預防和預警。
3. 系統技術：利用大數據、資料探勘和人工智慧相關的決策樹，將獲得之數據與資訊加以運用彙整，能提升預警的準確率。
4. 其他：透過此系統減少併發症的發生率，降低病患的傷害程度，有效提高醫療照護品質及病患安全。

現今我國人口高齡化狀況越來越嚴重，醫療費用的支出也日益沈重，藉由本系統

王乃弘、李曉慧、張肅婷、林政杰、林俊榮、王思怡 建構以大數據為基礎的併發症預警模型-以高血脂併發症為例

的建置，能使得醫護人員能獲得更多有效的資訊，也能經由併發症的預警讓病患獲得更好的照護與品質。透過疾病預測模型探討民眾是否為潛在病患，將確診前的病患的病歷資料整理分析過後，應用決策樹演算法得到有高血脂且有併發症前的特徵，做出預測系統，再透過系統進行比對推算，以獲得患病的機率，並能提供醫療院所和病患及早進行衛教和預防措施。

參考文獻

- 尹相志(2009)，SQL Server 2008 Data Mining 資料採礦(附 CD)。台北：悅知文化。
- 黃原博(2012)。應用資料探勘技術於慢性腎臟病之預測。臺東大學綠色科學學刊，2(1)，79-80。
- 蔡毓娟(2015)。展望大數據的應用-DDC 數位化論文典藏聯盟〔電子版〕。高大圖書資訊館通訊，47。
- 蔡詩怡(2011)，以探索性資料分析方法發展心臟血管疾病臨床輔助預知模型(碩士論文，國立台北護理護理健康大學，2011)。臺灣博碩士論文知識加值系統。
- 問上醫(2017)。高血脂的症狀及常見併發症。每日頭條。線上檢索日期：2018年11月27日。取自網址：<https://kknews.cc/zh-tw/health/mm5vq4p.html>
- 吳珮均(2017)。心肌梗塞來得突然!哪些症狀要小心?。健康醫療網。線上檢索日期：2018年11月30日。取自網址：
<https://www.healthnews.com.tw/blog/article/35440/%E5%BF%83%E8%82%8C%E6%A2%97%E5%A1%9E%E4%BE%86%E5%BE%97%E7%AA%81%E7%84%B6%EF%BC%81%E3%80%80%E5%93%AA%E4%BA%9B%E7%97%87%E7%8B%80%E8%A6%81%E5%B0%8F%E5%BF%83%EF%BC%9F>
- 彰化基督教醫院 血管醫學防治中心資訊網(2010)。高血脂。線上檢索日期：2018年12月10日。取自網址：
https://www.cch.org.tw/vmpc/knowledge/knowledge_2_detail.aspx?oid=102&no=7
- 衛生福利部統計處(2020年6月15日)。108年死因統計結果分析。衛生福利部統計處。取自網址：<https://dep.mohw.gov.tw/DOS/cp-4927-54466-113.html>
- 魯皓平(2017)。氣溫驟降慎防心肌梗塞復發!降低壞膽固醇保命!。遠見雜誌。線上檢索日期：2018年11月30日。取自網址：<https://www.gvm.com.tw/article/41047>