資料探勘技術應用於具周邊動脈疾病之透析病患之研究

吳淑梅[1,3]　羅家倫[1,2]　胡雅涵[1]
[1]國立中正大學資訊管理所　yahan.hu@mis.ccu.edu.tw
[2]嘉南藥理科技大學資訊管理系 allenlo.tw@gmail.com
3.財團法人天主教聖馬爾定醫院 wusu_may@yahoo.com.tw

摘要

根據美國腎臟資料登錄系統（US Renal Data System,2010,USRDS）公告資料顯示，台灣末期腎臟病患(End stage renal disease, ESRD)之發生率約每百萬人口 384 人，佔世界排名第三位，意即每壹百萬人口就有 384 位新發生必須接受腎臟替代治療的個案；盛行率約每百萬人口 2288 人，指的是每壹百萬人口中有 2288 人接受腎臟替代治療，無論是發生率或盛行率在美國腎臟資料登錄系統之排名皆高居全球第一位，高居世界排名第一位(USRDS,2010)。顯示腎臟病已是國人最嚴重的疾病之一。儼然已晉昇為新國病。

而末期腎臟疾病的病人比一般正常人更容易患有周邊動脈阻塞疾病(PAOD)，增加 PAOD 將增加住院率及死亡率。因此，若能早期知道那些末期腎臟病患，具有死亡的風險，對於醫療照護決策是否要對末期腎臟病患進行更積極性的治療評估有著極大的幫助。

本研究透過資料探勘的技術來建構本研究的預測模式，以提早了解具有 PAOD 疾病之透析患者疾病惡化的速度，我們同時也採用不同的監督式學習技術(supervised learning techniques)來進行實驗，並嘗試從中找出較佳實驗結果之預測模式，目的即為了提供一個不同的照護決策方式，以提高整體透析患者的醫療品質。

關鍵詞：資料探勘，末期腎臟病，決策樹，類神經網路

資料探勘技術應用於具周邊動脈疾病之透析病患之研究

1. 前言

根據美國腎臟資料登錄系統（US Renal Data System,2010,USRDS）公告資料顯示，台灣末期腎臟病患(End stage renal disease, ESRD)之發生率約每百萬人口 384 人，佔世界排名第三位，意即每壹百萬人口就有 384 位新發生必須接受腎臟替代治療的個案；盛行率約每百萬人口 2288 人，指的是每壹百萬人口中有 2288 人接受腎臟替代治療，無論是發生率或盛行率在美國腎臟資料登錄系統之排名皆高居全球第一位，高居世界排名第一位(USRDS,2010)(如圖 1)。顯示腎臟病已是國人最嚴重的疾病之一。

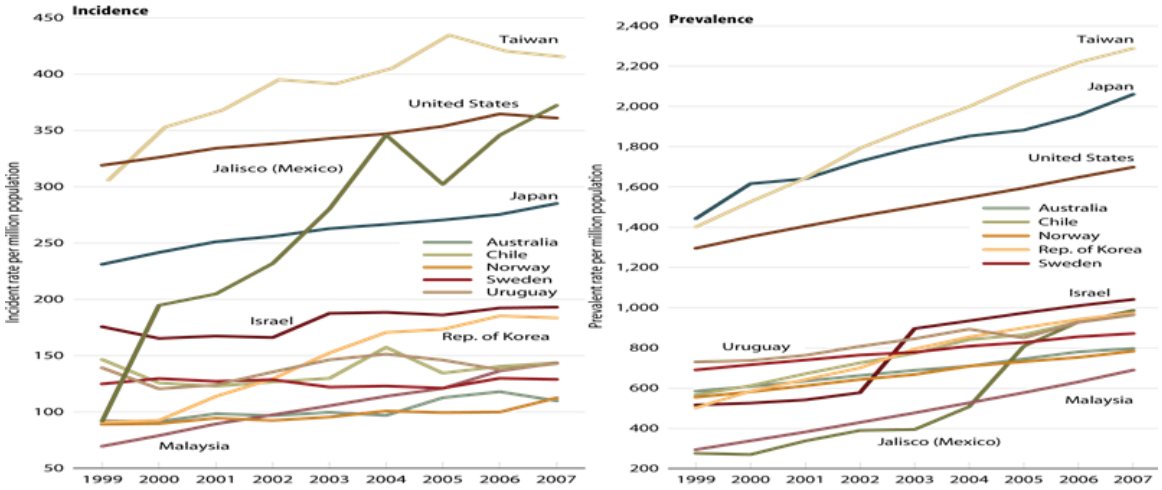

圖 1:世界 ESRD 盛行率與發生率比較圖

而末期腎病變需要接受長期透析治療或是換腎才能維持生命，由於腎臟來源得之不易，因此大部分罹患末期腎病變的病人選擇透析為其腎臟替代療法，其中又以血液透析居多(Guerrero et al., 1997；National Kidney Foundation, 2002)。我國自 1995 年實施全民健保(NHI)後，減少進入透析醫療的障礙，且國民平均餘命延長，死亡率持續減少，國內人口逐漸老化，65 歲以上末期腎臟疾病(ESRD)的盛行率及發生率更是急速的增加(圖 2)。根據台灣腎臟醫學會(2010)資料顯示，選擇血液透析治療的病人約 90.5%，選擇腹膜透析治療約 9.5%。截至 2011 年 6 月台灣領有「慢性腎衰竭(尿毒症)必須定期透析治療者」重大傷病卡有效領證數高達 67458 張,相較於 2010 年底又增加了 1575 張(行政院衛生署,2011)。

台灣末期腎臟疾病與慢性腎臟疾病(ESRD)高盛行率與高發生率之原因，除了人口老化，糖尿病、高血壓等慢性病崛起，治療進步降低死亡率卻使腎臟受損與失去功能的機會增加而導致續發性腎臟疾病外，根據國內學者研究推斷沒有及早轉診腎臟專科治療(Chen et al., 2008; Chen et al., 2010）、民眾對慢性腎臟疾病認知不足、高度使用中草藥如馬兜鈴酸(Kong.2008) 、濫用止痛藥（Wen et al.,2008；Guh etal.,2008）及其它具腎毒性的藥物如 Acetaminophen、Other NSAID、Aspirin 等都可能是末期腎臟疾病發生導致需接受腎臟替代治療的重要因素。

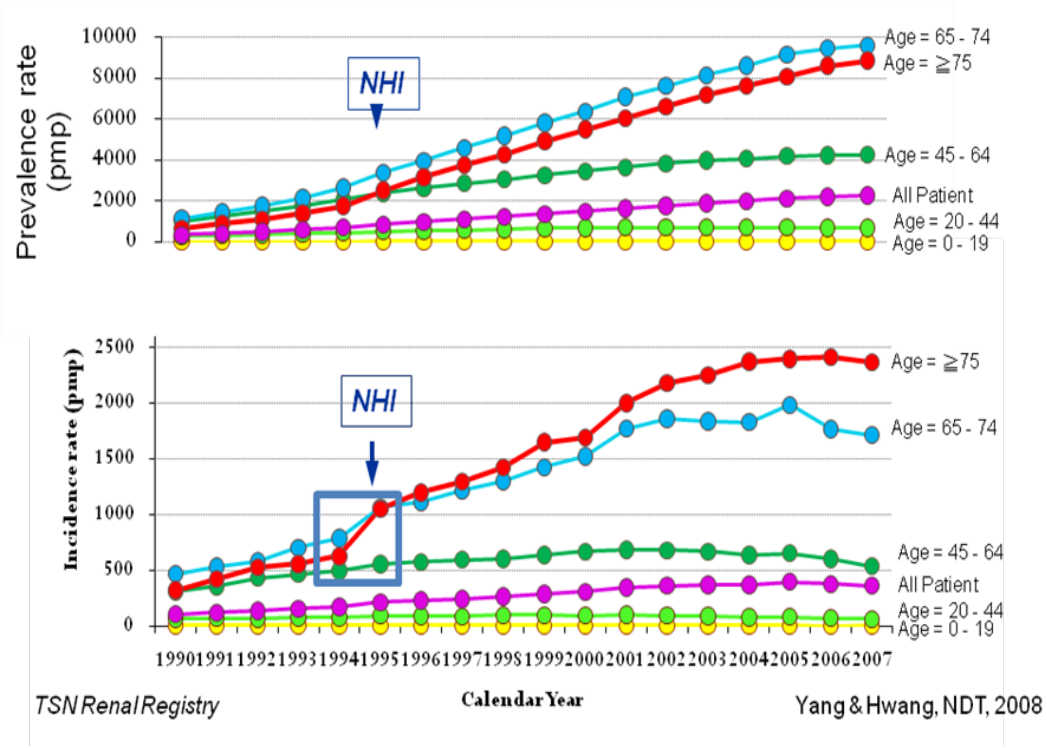

圖 2 :主辦單位：世界 ESRD 盛行率與發生率比較圖

根據衛生署統計，心臟疾病（高血壓性疾病除外）、腦血管疾病分別位於 2010 十大死因的第二及第三位。而末期腎臟病患者於臨床上常併有較高的心臟血管疾病，其中包含周邊動脈阻塞疾病（peripheral arterial occlusive disease; PAOD），而周邊動脈阻塞疾病是全身血管粥狀動脈硬化的表徵，隨著病程進展，動脈管徑越來越窄，影響血液灌流導致缺血症狀，與心血管疾病及死亡的發生密切相關（Newman et al., 1999）。

末期腎臟疾病的病人有周邊動脈阻塞疾病(peripheral arterial occlusive disease,PAOD)增加的風險，Selvin et al.(2004)研究發現一般族群罹患周邊動脈阻塞疾病的盛行率為 4.3% ，Cheung et al.(2000)、研究則指出血液透析病人同時合併有周邊動脈阻塞性疾病者卻高達 23% ，由上述研究可見末期腎臟疾病病人周邊動脈阻塞性疾病盛行率較一般族群高出數倍。Rajagopalan et al. (2006)研究也指出末期腎臟疾病病人併有周邊動脈阻塞性疾病的住院率及死亡率也比一般族群高。因此，對於末期腎臟病患來說，及早知道死亡的風險，對於醫療照護決策是否要對末期腎臟病患進行更積極性的治療評估有著極大的幫助，過去的相關研究中，極少數對於周邊動脈阻塞疾病和透析病患之間的關係，同時，研究樣本較少，同時亦只就單一醫療院所及單一等級醫院的個案進行探討，也因此可能造成偏頗的研究結果。

由於以上的原因，本研究希望透過資料探勘技術，並以中南部一家區域教學醫院、一家地區醫院、二家血液透析診所，周邊動脈阻塞疾病盛行率約 36%，遠遠高於國外及國內其他透析院所。病患共約 600 位透析病患中患者為對象，為對象，周邊動脈阻塞疾病偵測周邊動脈硬化疾病相關因子的大量的資料中，找出潛在有用的隱含特徵及易於理解的知識，透過周邊動脈硬化疾病相關變數與死亡結果之關係，建置出周邊動脈硬化疾

病引發死亡之預測模型，期望能提供各院區執行醫療服務時，可以早期評估及診斷並針對病患狀況提早預防或擬定診療計劃進行預先處置，減少死亡率並提升醫療照護品質。

2. 文獻回顧

2.1 末期腎病變及腎臟替代療法

根據美國腎臟基金會（National Kidney Foundation, NKF）於 2006 年發表的腎臟疾病預後之品質導引（Kidney Disease Outcomes Quality Initiative，K/DOQI）中將慢性腎臟疾病（Chronic kidney disease, CKD）第五期(末期)定義為腎絲球過濾率 (glomerular filtration rate; GFR)小於 15ml/min/1.73m^2(如表 1 所示)，是定義是指腎臟功能已達到漸進性、不可逆的嚴重損害。末期腎病變(ESRD)主要病因為代謝性病變：糖尿病、慢性腎絲球腎炎、慢性間質性腎炎：濫用止痛劑引起 、腎硬化症：高血壓導致、多囊腎：染色體(遺傳)等。可能出現的症狀及徵象為腎臟移除水分及維持恆定的功能下降，對身體的影響包括神經系統、腸胃系統、心臟血管系統、血液、代謝性酸中毒、腎性骨病變、體液電解質失去平衡、內分泌及皮膚的變化，症狀包括疲倦、嗜睡、周邊神經功能異常、噁心、嘔吐、貧血、出血、皮膚有搔癢感、水腫、血鉀過高、高血壓等。此時可能需要執行腎臟替代療法，以降低死亡率，減輕不適的症狀。

表1:CKD各分期的治療策略(美國國家腎臟基金會)

Stage	Description	GFR (ml/min/1.73 m^2)	Action
1	Kidney Damage with Normal or ↑ GFR	>90	診斷及治療 治療合併症 延緩腎功能惡化 減少心血管疾病危機
2	Mild ↓ GFR	60-89	預估腎功能衰退情形
3	Moderate ↓ GFR	30-59	評估及治療併發症
4	Severe ↓ GFR	15-29	準備腎臟替代療法
5	Kidney Failure	<15 or Dialysis	尿毒症出現時 開始替代療法

腎臟替代療法大致可分為血液透析、腹膜透析及腎臟移植，但由於找尋適當不會排斥病人的腎臟不是容易的事，因此大部份人選者前 2 者做為主要的治療方法根據 2009 年中央健康保險局資料顯示全國共有透析病人 59161 人，其中 53441 人選擇一周到醫療院所 2-3 次的血液透析治療作為其腎臟替代治療方式。透析人口 90.4%；另外 5720 人選擇居家腹膜透析治療，占透析人口 9.6%(台灣腎臟醫學會,2009)。，以下針對這 2 者做簡單的介紹：

血液透析治療是將人工腎臟用來替代腎臟移除廢物。其基本結構主要為 6000-15000 根半透膜的空心纖維，透析治療原理是利用人工腎臟空心纖維的半透膜利用其兩側的濃度差，將溶質會由高濃度往低濃度移動之原理，來幫助血中廢物的清除以及將體內需要的物質從透析液端擴散進入血液中並將血液中的毒素及多餘的水分經由擴散及超過濾的方式移除至體外，來暫時或永久的替代腎臟之排泄功能。長期血液透析病人由腎臟專科醫師依個別狀況評估所需治療之時間，週而復始每週 2-3 次，每次 3-5 小時到透析院所進行血液透析。

腹膜透析則是至 1978 年才被證明可行後正名，台灣地區則是至 1990 才納入勞保診療費用支付標準給付範圍。這種治療方式是以將一條矽質導管以手術的方式將一個包含臟膜層和壁膜層，覆蓋著腹膜腔的一個漿膜植入病患腹腔裡，以建立腹膜透析之透析液進出通路，於術後大約 10~14 天後可以開始將透析藥水經由導管灌入、利用肚子裡腹膜的半透膜性質進行擴散作用，使溶液中的溶質經腹膜由高濃度往低濃度移動，最後濃度呈現平衡狀態。水分的移除則利用透析藥水中高濃度葡萄糖的滲透壓進行超過濾作用，意即溶液中的水分子經腹膜由水分子容量多(滲透壓低)得地方往容量少(滲透壓高)的地方移動。體內聚積的尿毒素以及多餘的水分藉由擴散及滲透作用移至腹腔中，最後經由導管把這些含毒素及多餘水分的透析藥水從肚子引流出體外丟棄。

最近幾年發展出自動腹膜透析（Automated Peritoneal Dialysis, APD）的治療方式，病人於晚上睡前將導管與自動腹膜透析機連結上以後便可安心睡覺，機器本身會自動為病人進行透析液更換的步驟及監測。待隔天醒來將導管與機器分離，便可開始一天正常的活動。可以減少病人需固定時間回醫院進行透析治療的麻煩，同時由於導管銜接的次數減少，腹膜炎感染率也較為改善。不過，無論是執行 APD 或 CAPD 治療都還是必須由腎臟專科醫師依病人每月回診之各項檢驗、檢查報告、透析品質評估結果決定治療方式。治療劑量及換液次數依病人透析量需求之不同調整。

2.2 週邊動脈阻塞疾病(peripheral arterial occlusive disease; PAOD)

周邊動脈阻塞性疾病（peripheral arterial occlusion disease, PAOD）是主動脈弓以下的動脈產生粥狀硬化，隨著病程進展，動脈管徑越來越窄，影響血液灌流，造成其下肢的組織缺血症狀。根據研究統計指出周邊動脈阻塞性疾病在一般族群的盛行率只有 4.3%(Selvin et al., 2004)，但根據美國腎臟病資料登錄系統(USRD,2000)統計資料顯示：血液透析病患中卻有高達 23%比率會同時合併有周邊動脈疾病。國內學者 Lim et al.,(2005)針對某醫學中心 243 位長期透析患者的研究發現，長期透析患者罹患周邊動脈疾病的盛行率為 27.8%；國內另一篇由 Tsai et al.,(2008)的研究則指出，某醫學中心長期血液透析病人之周邊動脈阻塞疾病之盛行率為 24%。

末期腎病患者因此種疾病造成高度的心血管疾病住院率及死亡率及非外傷性截肢（Rajagopalan et al., 2006）。透析的病人經常會因周邊動脈阻塞疾病導致明顯的疾病惡化甚至死亡（陳美如&蔡世傑.,2008）。根據文獻指出，PAOD 的影響因子包括：年齡、家族史、抽菸、高血壓、高膽固醇、糖尿病、生活壓力壓力調適等可控制因素，血液黏滯度程度、低血清白蛋白濃度、低血鈣、高血磷等等是其影響因子。而要判斷是否患者 PAOD，一般而言，是以踝臂血壓比(ankle-brachial index, ABI)來進行判斷，踝臂血

壓比≥1.3 可代表有動脈鈣化的現象(Fishbane et al., 1995; Fowkes et al., 1991;Newman et al, 1997)。踝臂血壓比檢測方法為分別測量手臂及下肢之血壓，再以下肢的收縮壓除以上肢的收縮壓，所得之數值即是。當數值越小表阻塞的程度越厲害，症狀也更嚴重。數值 1.0 左右是正常範圍。介於 0.9-0.7 屬輕度的阻塞，會間歇性跛行。介於 0.7-0.4 是中重度阻塞，症狀為缺血性疼痛。等於或小於 0.4 則是為重度阻塞，肢體可能會壞死。

綜合以上文獻所述可以得知，雖然臨床上知道 PAOD 疾病所導致對於透析患者死亡造成的很大的間接影響，但卻少有相關研究以 PAOD 來建構透析患者的預測模型，以提早針對這些患者進行監控，可以早期評估及診斷並針對病患狀況提早預防或擬定診療計劃進行預先處置，減少死亡率並提升醫療照護品質。

3. 方法

Prediction analysis 的目的是透過監控式學習（supervised learning techniques）來建構一預測模式，以找出物件集合中 input variable 與 output variable 之間的關係。首先，Prediction analysis 針對一群已知變數屬性數值的物件來進行運算，並建構預測模式，最後再將未知 output variable 或是新的物件帶入預測模式，以進行預測。

3.1 C4.5 決策樹(J48)

C4.5 是為了改善 ID3 演算法易產生過度配適(Overfitting)的問題，而發展出來的決策樹歸納學習法(Quinlan, 1986)，C4.5 會選取有最大 GainRatio 的分割變數作為準則來加以改進，是目前最常使用的決策樹分類技術。ID3 主要是利用資訊獲利(Information Gain)的計算方法，對各屬性進行運算，取資訊獲利較大的屬性。而資訊獲利是決策樹中常用來做為特徵選取的標準，通常會選擇資訊量最高的屬性做測試。計算各類別資訊量，並進而訓練集合的平均資訊量，稱之為亂度也就是熵值(Entropy)。熵值通常用來表示資料的複雜度。但 ID3 在處理一般分類的問題上已有不錯的成效，但是若分割出來最後子集合中都各只有一個資料時，其分割後得到最大的資訊量，其值為零，在這種情況下，分割將沒有太大的意義，為了改善這樣的問題（也就是過度配適）Quinlan 在 C4.5 中提出將 Gain 正規化，當子集合個數越多時正規化後的值就會越大，GainRatio 相對會偏小，因而有效改善了 ID3 所容易產生的過度適配的問題。

3.2 類神經網路(ANN)

Artificial Neural Network (ANN)是一種模擬生物神經網路的數學模式(mathematical model)，利用系統輸入與輸出所組成的資料以建立輸入與輸出間的關係。在各種類神經網路學習模式中，back propagation neural network (BPNN)是目前最具代表性、應用最廣泛的一種supervised learning network (Rumelhart Hinton, Williams, 1986)，屬於多層的網路結構，包含input layer、 hidden layer，以及output layer。輸入層由問題之輸入變數所構成，此層的節點會接受輸入資料，然後傳遞到隱藏層進行計算。隱藏層位於輸入及輸出層之間，為整個類神經網路計算的主要部分，包含合併函數(combination function)及轉換函數(transfer function)，用以表現輸入單元之間的交互影響，此層的節點數目並無規則可循，通常需要經過試誤(trial and error)才能決定。輸出層為類神經網路的末端，主要用於輸出網路計算的結果。

BPNN 透過網路輸入與目的輸出，經過運算產生實際輸出，且不斷地比較目的輸出與實際輸出，而加以調整神經元之間的權重值，經過反覆的學習，使網路能得到更為接近目標值的輸出，來達成學習的目的。在 BPNN 的學習過程中，網路架構與參數的調整是重要的。在網路學習參數方面，有兩個參數必須做適切的調整：learning rate 與 momentum。learning rate 會影響網路的收斂效果，其值通常介於 0~1 之間。

3.3 AdaBoost 演算法

Adaboost(Adaptive Boosting)是由Yoav Freund 與Robert Schapire(Freund ＆Schapire, 1999)在1995年提出的一種機器學習演算法，他是以傳統的委員會機器為基礎，再加入順序及決策權重的觀念，提供AdaBoost委員會機器中的各個分類器不同的權重。其主要精神在於針對每個分類器，對學習不好的訓練資料範例再加重學習，因此Adaboost所建構出的多個分類器有前後順序之分。圖3中顯示AdaBoost委員會機器訓練流程，於第一回合剛開始時，資料集合中所有資料的權重一律相等，但是經過每個回合的訓練階段後，被錯誤分類的樣本權重將會增加，在資料驗證並計算正確率時，重新給予每個訓練資料點新的權重，因此在前一個分類器中被分類錯誤的資料點將會因此取得較大的權重，當下次再做分類器建構時，將會以新的權重來產生新的分類器，在每個回合中，分類器將被迫面對前一個分類錯誤的範例而進行加強學習，以便修正缺失。如此週而復始的循環，最後，將建構出Ada-Boost委員會機器，其決策結果將考慮每個分類器的權重，以加權平均的方式得到最後的分類結果。

4. 實驗

4.1 資料收集與變數定義

本研究在確定問題後，經過去的相關研究結果與專業醫師密切討論後，透過與雲嘉南地區透析中心合作(分別包含一家區域醫院、一家地區醫院及二家診所層級)來取得研究資料，分別在這些院所於 2008 年 3 月至 6 月間找出曾接受 ABI 檢測的透析患者共 543 位，進而分析其 ABI 檢測結果後，找出符合周邊透脈疾病患者共 182 位，由於每位透析患者在治療期間皆需於每季固定量測許多檢驗查查資料，故本研究根據文獻至醫院中之 HIS 及 LIS 資料庫收集於該時間內最近但不大於該時間之 13 項基本資料及衡量變數，分別為性別、年齡、糖尿病、高血壓、身體質量、白血球、血球容積、白蛋白、總膽固醇、三酸甘油脂、空腹血糖、轉鐵蛋白、B 及 C 型肝炎。182 位患者中，截至 2011 年 6 月 30 日止，已死亡個案為 66 例。而本研究是為了建立死亡的預測模式，因此以死亡與否為應變數。

4.2 資料前處理

資料收集後，同時和腎臟科醫師討論後將，依取得之資料表欄位修正、刪除，差異過大的離群值則由專業腎臟科醫師判斷是否刪除，同時若有一些不合理或差異過大之離群值不具分析價值之欄位刪除，以求進行探勘時不因為多餘之資料欄位而影響到最終結果之準確性。其次，為使欄位格式符合演算法的要求，本研究也將原始欄位進行轉換，

如生日將其轉換為年齡、血液生化資料表內屬連續性變數的檢驗值以合理的方式予以切割，以轉換為名目尺度。個案基本資料表欄位取病人 ID、病歷號碼、年齡、性別、此四個欄位，其餘如住址、個人電話、學歷、婚姻等欄位因非患者疾病的死亡之因子故將予以刪除。另就血液生化檢驗資料表因常規透析病患檢驗檢查頻率、時間、項目有標準規範，故腎臟醫學會所開發的軟體中紀錄了大量、完整且有時序性的檢驗資料值，但在紀錄中之部份欄位因病患住院或出遊請假等因素造成有些欄位會出現遺漏值之情形。經資料表欄位簡化修正，最後，分別以各別醫院的病歷號碼為主欄位合併分別取出之各項資料，再將 4 家醫院合併成一個實驗用的關聯表資料集。並進行處理、淨化後之資料整合成供實驗之資料表。

4.3 實驗及評估方式

本研究採用 WEKA3.7.3 版開放原始碼應用軟體之 J48 及 MLP 模組作為本研究 Warfarin 用藥劑量模式樹的建構與分析工具，本研究實驗資料集由於應變數分佈不均，為求實驗結果不致偏頗，除採用 10 摺交叉驗證法(10-fold cross-validation) (Kretschmann, Apweiler, 2001)進行各組預測模式的效能評估，以克服只抽樣一次所造成的資料不平均，而影響分類器預測的準確性。也將參數的 random Seed 之設定採用 1 至 10 來產生 10 組不同的子集合，再將 10 次的實驗結果的模式效能加以平均後用以評估各種分類器的效能優劣。其次，本研究期望建構出 Warfarin 初始劑量的預測模式，並透過不同分類技術，比較各種預測模式之預測正確率，最後再與原始資料之正確率進行效能的比較。

5. 實驗結果與討論

由於決定是否有週邊動脈疾病之踝臂血壓比在盛行率高的透析病人中並非常規檢查，故腎臟科醫師通常藉由病人相關徵候或症狀之主訴及四肢皮膚評估來判斷周邊動脈阻塞疾病，且須將病人轉介心臟科確診及進一步處置。但是超過二分之一的周邊動脈阻塞疾病病人沒有明顯的臨床症狀(Hirsch et al., 2001)，造成臨床之診斷十分的困難，進而導致治療延誤，嚴重造成危急病人生命的後果，本研究透過資料探勘技術來建構病人最終會惡化至死之具周邊動脈疾病透析病人預測模式，以實際個案最終於三年內因疾病惡化而死亡的個案當做基準值(baseline)，比較經決策樹和 ANN 二種分類器來建立預測系統後之預測效能，再分別以 Adaboost 分類提昇技術來改善預測效能，進而決定較好的預測系統

5.1 單一分類器

首先，在單一分類器的實驗結果中（如表二所示），C4.5 與 ANN 的十次實驗的平均正確率分別為 79.12%、85.16，相較於原始資料集中之最後資料分佈的死亡比率準確率大幅增加了 43.12%和 49.16%。同時，以 2 個實驗之十次的標準差($\sigma(E)$)來觀察預測模式的穩定度，也可以發現無論以 J48 或 ANNSVR 或 ANN 所建構出來的預測分類器，穩定度也都十分的高。若比較二個分類器的實際結果，以 ANN 所建構的患者死亡預測模式之準確率比以 C4.5 所建構的預測模式準確，改善幅度約為 6.04%。因此，在本研究中 ANN 是單一分類器準確率較佳的分類技術

表二:單一分類器實驗結果

單一分類器	正確率	σ(E) of 10 seed
Baseline	36%	-
C4.5(J48)	79.12%	0.07
ANN(MLP)	85.16%	0.04

5.2 多重分類器

其次,在以 Adaboost 及 Vote 結合原有單一分類器的實驗結果中顯示(如表三所示):二個分類器透過 Ada-Boost 進行實驗結果效能提昇後,確有改善,不過分別都些微提昇了 0.55 至 79.67%及 85.71%,而以 2 個實驗之十次的標準差($\sigma(E)$)來觀察預測模式的穩定度的結果,和單一分類器時的差異不大,但若以 Vote 結合全部的分類技術之多重分類器的實驗結果,分類效能大幅提昇 91.81%,相較於已經透過 Adaboost 進行效能提昇之 Ada+J48 及 Ada+MLP 分類結果,還再改善了 12.12%及 6.09%,顯示以 Vote 再結合以 Adaboost 改善後的 2 個分類器為最佳的分類技術。

表三:多重分類器實驗結果

多重分類器	正確率	σ(E)
Ada+J48	79.67%	0.06
Ada+ MLP	85.71%	0.04
Vote(Ada+J48& Ada+MLP)	91.81%	0.06

5.3 決策樹規則分析

雖然本研究的實驗結果,以 VOTE 實驗結果較佳,但由於 MLP 無法透過明確的規則分析來進行細部觀察以進一步了解具周邊動脈之透析患者病程較容易惡化之特微屬性,因此本研究選擇以 J48 建構之決策樹分類模式取出準確率較高且較明確的四項規則如下供參考:(一)預測模式 1:Albumin >3.5、Ferritin<100、有肝炎、三酸甘油脂高於 150,罹患周邊動脈疾病大於 65 歲的女性透析患者,有較高死亡率的現象。(二)預測模式 2:Albumin >3.5,沒有罹患肝炎但有高血壓病史的 65 歲以上罹患周邊動脈疾病的女性透析患者,有較高死亡率的現象。(三)預測模式 3:Albumin >3.5,沒有罹患肝炎的 65 歲以上罹患周邊動脈疾病女性透析患者,總膽固醇>150mg/dl 有較高死亡率的現象。(四)預測模式 4:Albumin <3.5,膽固醇介於 150-300mg/dl,血球容積<26%;轉鐵蛋白高於 500ng/ml;罹患周邊動脈疾病的透析患者,有較高死亡率的現象。

6. 結論

我國洗腎患者無論發生率或盛行率在世界中都是數一數二的,而洗腎患者由於同

時患有周邊疾病造成快速死亡的機率又比其他疾病要高上去多，因此，本研究期待建立一個具有周邊疾病之末期腎臟病患者會因為疾病惡化最後導致死亡之預測模式，讓臨床上能及時運用以掌握這些關鍵病人以改善透析患者的醫療品質。本研究採用 C4.5 及 ANN 來建立患者死亡的的分類模式，目的即為了據之以改善目前未能確實掌握之具周邊患者疾病之透析患者的照護模式。本研究實驗結果證明，以資料探勘技術所建構的預測模式，可以有效預測惡化的機率值

未來的研究方向上，本研究雖然收集了各層級的醫院，但未能分別以權屬別來比較是否不同層級院所之預測模式會有所不同。同時，也可擴大到跨區域的聯盟醫院體系，可以讓本研究之預測模式更經得起外推性的考驗。最後，也可以較血液透析和腹膜透析患者是否會有所不同的實驗結果。

7. 誌謝

本研究係由國科會專題研究計畫補助(計畫編號：NSC100-2410-H-194-024-MY2)。

參考文獻

1. 洪國騰、吳明修(2003)‧末期腎病之下肢動脈阻塞性疾病‧腎臟與透析，15(1)，41-44。
2. 陳美如、蔡世傑、陳宣志(2006)‧周邊動脈阻塞疾病‧基層醫學，21(11)，318-325。
3. 陳杏婉（2008）．血液透析患者周邊動脈阻塞性疾病之篩檢及相關因素分析．未出版的碩士論文，桃園縣：長庚大學護理學研究所。
4. 賴怡青、曾春典（2007）．周邊動脈疾病的內科治療，臺灣醫界，50（2），6－10。
5. Cheung, A. K., Sarnak, M. J., Yan, G., Dwyer, J. T., Heyka, R. J., Rocco, M. V., Teehan, B. P., & .Levey, A. S. (2000). Atherosclerotic cardiovascular disease risks in chronic hemodialysis patients. Kidney International, 58**,** 353-362.
6. Fishbane, S., Youn, S., Kowalski, E. J., & Frei, G. L. (1995). Ankle-arm blood pressure index as amarker for atherosclerotic vascular diseases in hemodialysis patients. American Journal ofKidney Disease, 25**,** 34-39.
7. Fowkes, F. G., Housley, E., Cawood, E. H., Macintyre, C. C., Ruckley, C. V., & Prescott, R. J.(1991). Edinburgh Artery Study: prevalence of asymptomatic and symptomatic peripheral arterial disease in the general population. International Journal of Epidemiology, 20**,** 384-392.
8. Elizabeth Selvin, Spyridon Marinopoulos, Gail Berkenblit , Tejal Rami, Frederick L. Brancati, Neil R. Powe, and Sherita Hill Golden, Meta-Analysis: Glycosylated Hemoglobin and Cardiovascular Disease in Diabetes Mellitus; Annals of Internal Medicine; 2004, vol. 141 no. 6 421-431
9. Freund, Y., & Schapire, R. E. (1999). A Short Introduction toBoosting. *Journal of Japanese Society for Artificial Intelligence, 14*, 771–780.
10. 23.Guerrero, A., Montes, R., Muñoz-Terol, J., Gil-Peralta, A., Toro, J., Naranjo, M., et al.(2006).Peripheral arterial disease in patients with stages IV and V chronic renal failure.

NephrologyDialysis Transplantion, 21, 3525-31.

11. Hsieh, L. Gandour, J., Wong, D., & Hutchins, G. D. (2001).Functional heterogeneity of inferior frontal gyrus isshaped by linguistic experience. Brain and Language, 76,227 –252.
12. Kretschmann, E., and Apweiler, R. “Automatic rule generation for protein annotation with the C4.5 data-mining algorithm applied on peptides in Ensembl” Proceedings of the German Conference on Bioinformatics Braunschweig, Germany 2001, pp:53-57.
13. National Kidney Foundation, ; Clinical Practice Guidelines for chronic kidney disease: evaluation, classification and stratification. Am J Kidney Dis. 2002; 39:S1-266
14. Newman, A. B., Tyrrell, K. S., & Kuller, L. H. (1997). Mortality over four years in SHEP participants with a low ankle-arm index. Journal Of The American Geriatrics Society, 45,1472-1478.
15. Rajagopalan et al., 2006 R. Rajagopalan, H. Vaucheret, J. Trejo and D.P. Bartel, A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. *Genes Dev.*, **20** (2006), pp. 3407–3425.
16. Rumelhart, D.E., Hinton, G.E., and Williams, R.J. “Learning representations by back-propagating errors” Nature(323) 1986, pp:533-536.
17. SZU-CHIA CHEN, JER-MING CHANG, MING-CHIN CHOU, MING-YEN LIN, JUI-HSIN CHEN, JIA-HUI SUN, JINN-YUH GUH3,SHANG-JYH HWANG, HUNG-CHUN CHEN, Slowing renal function decline in chronic kidney disease patients after nephrology referral, Nephrology; (2008)Volume 13, Issue 8, pages 730–736,
18. Tze-Wah Kao 1, 2, Jenq-Wen Huang 1, Kuan-Yu Hung 1,Yu-Yin Chang 2, au-Chung Chen 2, 3, Chung-Jen Yen 1,Yung-Ming Chen 1, Tzong-Shinn Chu 1, Ming-Shiou Wu 1,Tun-Jun Tsai 1, Kwan-Dun Wu 1, Jung-Der Wang; Life expectancy, expected years of life lost and survival of hemodialysis and peritoneal dialysis patients,journal of Nephrol,2010, 23(06): 677-682
19. U S Renal Data System, USRDS 2010 Annual Data Report:Atlas of Chronic Kidney Disease and End-Stage Renal Diseasein the United States, National Institutes of Health, NationalInstitute of Diabetes and Digestive and Kidney Diseases,Bethesda MD, 2010

Application of Data Mining Techniques to Constructing the Prediction Model for the Survival of Hemodialysis Patients

Su-May Wu[1] Chia-Lun, Lo[1,2] Ya-Han Hu[1]
[1]National Chung Cheng University yahan.hu@mis.ccu.edu.tw
Chia Nan University of Pharmacy & Science allenlo.tw@gmail.com
[3]St. MARTIN DE PORRES Hospital wusu_may@yahoo.com.tw

Abstract

According to the statistics form USRD, the incidence and prevalence of hemodialysis diseases is now ranked the first highest in the world as the number of patients keeps increasing and the dialysis population in Taiwan is now up to over 40,000 people.

End stage renal disease (ESRD) patients with POAD are increasing the risk of admission and dead. Therefore, if we discover the symptom and put it into remedy early, we can stop the end stage renal disease from rising, and then improve the early onset of end stage renal disease so as to reduce the waste of health care recourses.

This study refers to the cases of chronic renal disease patients managed by a southern regional hospital. We try to examine the supervised learning techniques that we select to construct the prediction systems and find out a trustworthy prediction model for the survival of hemodialysis patients. Finally, the experimental outcome indicates the well performance of the prediction system.