

脂肪肝檢測系統開發及績效衡量

傅家啟
雲林科技大學
工業工程與管理(系)所 教授
fujc@yuntech.edu.tw

周盈君
雲林科技大學
全球運籌管理所 研究生
g9626702@yuntech.edu.tw

詹德信
台灣歐瑞康萊寶真空(股)公司
經理
eric.chan@oerlikon.com

摘要

現代人隨著社會型態與飲食習慣的改變，容易引發許多慢性的疾病，而脂肪肝就是其中一個例子。通常脂肪肝沒有明顯症狀，所以很多人容易疏忽，直到身體出現一些不適症狀或健康檢查時，才得知罹患脂肪肝。

本論文將中部某醫學中心所提供之 1133 筆健檢資料，以亂數分類方式分為訓練組 568 筆，測試組 565 筆。首先使用訓練組資料，建立(1)全部特徵+ General Regression Neural Networks (GRNN)；(2) Sequential Forward (SFS) 特徵選擇+GRNN；(3)分類與迴歸樹(CART)等三種分類器。

以測試組資料驗證，實驗結果顯示利用 SFS 特徵選擇建構於 GRNN 分類器，在此分類器，其正確率為 0.7416；而採用全部特徵所建構的 GRNN 分類器及 CART 分類樹之正確率分別為 0.6867 與 0.6938。所以整體而言，GRNN 分類器結合 SFS 特徵萃取之分類績效最佳。

關鍵詞：脂肪肝、資料探勘、分類器

Abstract

With the modern life style and eating habits, it is prone to lead to many types of chronic diseases. Fatty liver is one of them. Since fatty liver does not have obvious symptom, it is often not discovered unless patients conduct the health check. Fatty liver is the collection of excessive amounts of fats inside liver cells. As long as the fat in the liver organ higher than 5% or the fat vacuoles found in the liver biopsy higher than 10%, it can be diagnosed as fatty liver.

In this paper, the health check of 1133 cases from a central medical center is applied to conduct the test. The data are randomly divided into the training data set (568 cases) and the test data set (565 cases). The training data set is used to develop a CART classification tree and GRNN without feature selection and with 28 selected features by Sequential Forward Selection. The test data set is used for performance evaluation of

the CART and GRNN.

Experimental results showed that the accuracy of the GRNN classifier applied the input features selected by SFS was 0.7416. The accuracies of the GRNN with all features and the classifier CART were 0.6867 and 0.6938, respectively. On the whole, GRNN classifier combined the SFS feature selection provides the best classification performance.

Key words: fatty liver, data mining, the classifier

1. 緒論

正常人肝組織中含有少量脂肪如三酸甘油酯、磷脂、醣脂以及膽固醇，其重量約為肝重量的 4~5%。脂肪肝主要是因為三酸甘油酯代謝異常，使得肝細胞質囤積過多的脂肪顆粒，而影響細胞質本身的功能，導致肝機能異常所致。因此，肝內脂肪含量若超過肝重量的 5% 以上，或者肝組織切片超過 10% 以上的肝細胞有脂肪空泡堆積的情形時，可診斷為罹患「脂肪肝」。

常見的原因為肥胖、糖尿病、營養不良、缺乏運動及高血脂肪、長期酗酒、C 型肝炎、藥物等，而絕大部份輕度脂肪肝都不具有臨床症狀，只有少數個案出現食欲不振、全身倦怠或腹脹，所以大多數人是在無意間做了肝臟超音波檢查才認識到脂肪肝的診斷。其中，非酗酒引起之脂肪肝的臨床意義，主要在於經由脂肪肝的發現，而發現肥胖及高血脂等致病因素，所以能早期治療並減少併發症的發生。

而脂肪肝病通常不容易察覺，需由專業醫師進行腹部超音波檢查才可發現(姚志成，2005)【4】，但目前中央健保局的成人健康檢查項目，大致分為四大部分：一是理學檢查、二是危險因素評估、三是血液檢查、四是尿液篩檢，且因為要考慮成本支出與該項疾病的盛行率，所以並未將腹部超音波檢查納入檢查項目(賴世偉、廖光福、黃金財，2005)【1】。因此，本論文目的是期望從醫院健檢中心資料庫掘取相關健康檢查資料，接著結合資料探勘(Data Mining)中之 SFS 為主要特徵擷取方法，

最後利用 CART 及 GRNN 兩種不同分類器進行分類，進而選取分類績效較好的分類器，為預防醫學領域中的脂肪肝防治工作，提供一醫療決策支援工具，然後針對可能患者進行腹部超音波檢測，使醫療資源有效的發揮及降低醫療成本。

健檢資料方面，主要是以：性別、年齡、身高、體重、脈搏、收縮壓(坐姿)、舒張壓(坐姿)、脂肪率、身體質量指數、白血球、紅血球、血色素、血球容積比、紅血球體積、平均血紅素、平均血球血紅素濃度、血小板數、尿比重、尿酸鹼值、尿膽元、結晶、甲型胎兒球蛋白、癌胎抗原指數、同半胱胺酸、空腹血糖、糖化血色素、血清蛋白、白蛋白、直接膽紅素、全膽紅素、GOP(肝功能)、GPT(肝功能)、rGT、鹼性磷酸酶、膽固醇、高密度膽固醇-脂蛋白、低密度膽固醇--脂蛋白、三酸甘油脂、尿酸、尿素氮(腎功能)、肌酸酐(腎功能)胰澱粉酶、鐵蛋白、肝膽胰脾臟超音波等 43 項，作為本論文分析項目。而且本論文是針對台灣中部某醫學中心接受全身健檢之資料庫進行資料探勘，所以不另外進行資料抽樣問卷調查及實施醫學臨床實驗測試。

此外，本論文限制方面，可列為以下三項：

1. 由於樣本對象為教學醫院健康檢查之一般民眾，因此屬於只收集一次的健康檢查相關資料，但某些健檢變數會隨著時間和身體情況的改變而有所不同，故不在本論文的探討範圍中。
2. 許多健檢變數無法事先規劃而缺少，如飲食習慣、運動習慣以及對脂肪肝的認知等，因此相關的個人習慣變數將不探討。
3. 資料探勘分析的結果，無法取代客觀的臨床醫學實驗數據，但是可以提供實質分析結果提供醫院參考。

2. 文獻探討

本論文是將健檢資料以探勘分類方式，經由 SFS 特徵擷取以及全部特徵擷取建構於 GRNN 分類器上，然後將健檢資料中的測試組，以 GRNN 分類器和 CART 分類樹分析關鍵特徵，診斷是否為脂肪肝。所以相關文獻包含脂肪肝潛在危險因子探討、特徵組合資料探勘、廣域迴歸神經網路(GRNN)、分類與迴歸樹(CART)等。

2.1 脂肪肝潛在危險因子探討

危險因子主要有肥胖、體脂肪率、肝指數等。

2.1.1 肥胖

脂肪肝形成之主要因素大致上可分為肥胖、高血脂症、糖尿病控制不良、飲酒、C 型肝炎及藥物反應等因素引起，其中肥胖、高血脂症、糖尿病控制不良等因素皆與肥胖有密不可分之關係(廖運範、陳東榮，1992)【2】，而根據統計，肥胖所形成的脂肪肝所需要的時間大約在五年以上。

通常肥胖性脂肪肝罹患者口服定量的葡萄糖後，血清胰島素的濃度高達正常的三倍以上；而單純肥胖者僅比正常稍高一些，由此可知，高胰島素血症是肥胖性脂肪肝特有表徵。此類肥胖性脂肪肝者，肝臟製造脂蛋白的量也比其他個體增加，所以會造成血清三酸甘油脂過高。

此外，脂肪變性的發生率及其程度，與個體肥胖的程度有一致正向的結果。事實上，重度肥胖在長期追蹤中，某些患者的肝臟有可能會發生纖維變性，甚至脂肪肝中細胞產生壞死現象是發生在某些女性而非酗酒者。因此，女性肥胖者對於飲食熱量的攝取，以及體重的控制是非常需要的。

2.1.2 體脂肪率(Percentage Body Fat, PBF)

所謂「體脂肪率」，是指「人體脂肪」與「體重」之百分比。以往判斷個人胖瘦時，通常是以身高及體重之比率(即 BMI，身高除以體重的平方值)來判定，往往忽略了人體脂肪之多寡，但像是從事健美運動的男子，其 BMI 可能很高，但體脂肪率卻不高，原因極有可能是體內瘦肉組織(LEAN BODY MASS)較多所致，所以要判斷一個人真正的肥胖程度，還是要檢測他的體脂肪率，才會比較客觀。

針對「體脂肪率」與「健康指數」的因果關係，有研究顯示體脂肪率與高密度脂蛋白膽固醇濃度呈顯著的負相關，而與男性的低密度脂蛋白膽固醇、三酸甘油脂、空腹血糖及女性的血壓呈顯著正相關，證明了「控制體脂肪」的重要性！因此，在評估肥胖方面，體脂肪率是真正客觀表現一個人脂肪含量的指標，也比其他方式準確(Tanaka etc., 2002)【10】。

2.1.3 肝指數指標

包含轉胺酶(AST, 簡稱 GOT), 麩氨酸丙酮酸轉胺酶(ALT, 簡稱 GPT), 鹼性磷酸酶(alkaline phosphatase, Alk-P), 膽紅素, r-麥胺基轉化酶(r-guutamyl transpeptidase, r-GT)白蛋白及球蛋白等。

其中鹼性磷酸酶是膽囊與肝臟相連的膽管內所分泌的一種酵素, 它的數量可以反映肝臟的問題, 因此, 當它的數目偏高, 就有可能是膽管出現了問題如阻塞、肝硬化甚至肝癌了。而膽紅素是紅血球內血紅素的代謝物, 經肝細胞處理後變成膽汁, 所以當肝臟細胞受損而無法處理膽紅素, 或膽管阻塞使膽汁無法順利排出時, 膽紅素就會逆流至血液中。所以膽紅素值上升, 有機會是膽管炎或是閉塞、肝細胞受損、肝炎或肝硬化的病徵。

2.2 特徵組合資料探勘

在國內外資料探勘領域中, 它除了用在顧客關係管理的部份很多外, 在醫學上資料探勘的實用例子也不少。國外部份: 例如 Demsar et al. 提出藉助決策樹的技術找尋重度創傷患者主要特徵, 並研究指出患者在加護病房所測的 pH 值與部份凝血激素時間為主要的因子, Demsar et al.(2001)【7】。Ramirez et al. 將資料探勘應用在愛滋病患者特徵之研究, 其先以 C4.5 決策樹篩選變數後再配合類神經網路辨識愛滋病患者, Ramirez et al.(2000)【8】。

資料探勘最主要的目的為經由特徵萃取得到特徵, 以及如何在所有擷取出的特徵中, 選擇對於肝脂肪分類有顯著貢獻之特徵組合。而不同的特徵變數之間, 不一定能確定是否有重複性或是相依性, 因此必須懂得如何適度的、有技巧的掌控與操作這些變數。尤其在分類問題上, 透過特徵擷取不但可以過濾不必要的特徵變數, 以簡化問題的複雜度, 還能在固定維度的組合情形下, 進一步來改善問題的正确率(楊敦翔, 2004)【5】。而本論文所使用的技術為逐次前饋式搜尋法(Sequential Forward Selection, SFS)。

2.2.1 逐次前饋式搜尋法(Sequential Forward Selection, SFS)

逐次前饋式搜尋法(Sequential Forward Selection, SFS)是由一維度的特徵開始逐一測試, 選擇出能使系統分類的誤差降至最小的特徵組合, 然後依據特徵的重要性, 開始一次增加一個維度特徵, 直到符合所要求的特徵變數

為止。

Yu(2000)【11】曾在乳房 X 光影像檢測時, 使用 SFS 搭配分類器做特徵選擇。SFS 分類誤差的評量是以均方差(Mean Square Error)作為依據, 定義如公式(1):

$$E = \frac{1}{q} \sum_{i=1}^q (d_i - a_i)^2 \quad (1)$$

在此 i 是第 i 個被識別的資料, a_i 是分類器辨識結果, d_i 是實際結果, q 是所有的訓練樣本數(蔡明倫, 2002)【3】。接著從訓練組(Training Data Set)中計算出所有特徵, 以 SFS 的方式產生出特徵子集合, 概述如下: 假設一開始之特徵有 {A、B、C} 三個特徵,

1. 先個別計算特徵子集合 {A}、{B}、{C}, 找出單一維度最具代表的特徵, 即均方差最小者, 假設為 {C};
2. 再依照上述步驟, 計算二維度特徵子集合 {C A}、{C B} 之均方差, 找出最小者, 假設為 {C A};
3. 最後再增加一個維度, 計算特徵子集合 {C A B} 之均方差。

所以我們得到經由 SFS 而產生的特徵子集合 {C}、{C A}、{C A B}, 假設其中均方差最小者為 {C A}, 則該特徵子集合 {C A} 即為 SFS 架構下之最小均方差特徵組合(王元宏, 2004)【6】, 示意圖如圖 1 所示。

2.3 廣域迴歸神經網路(General Regression Neural Network, GRNN)

近年被廣泛利用的類神經網路預測方法是藉由自我學習的概念, 去適應預測環境的改變, 並利用訓練樣本提供模型來連結輸入變數與輸出間的關係, 具有接受邏輯、數值、有序分類、無序分類變數作輸入、適應性強等特性, 因此, 類神經網路的優點就是「模型建構能力強」。

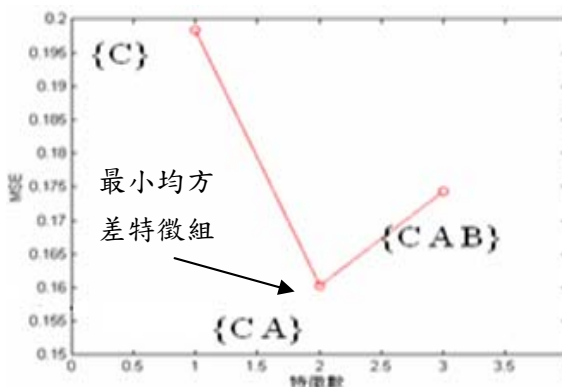


圖 1 SFS 特徵選擇示意圖

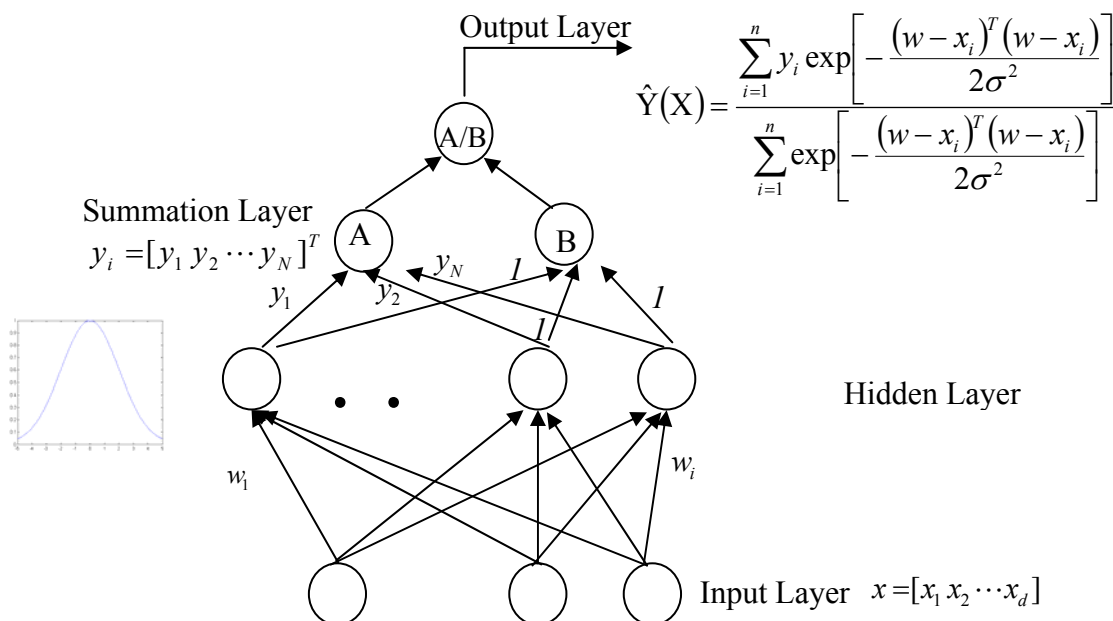


圖 2 GRNN 類神經網路架構圖

本文所提到的廣域迴歸神經網路(GRNN)是從 PNN(Probability Neural Network, 機率類神經網路)所演變而來,為監督式學習網路的一種。其中 PNN 是由 Specht (1988) 提出,但是 PNN 只適用於分類問題,而無法解決連續變數問題;Specht (1991)【9】又提出了 GRNN 的學習演算法,GRNN 可學習一個動態模式作為預測或控制用,因此迴歸問題不論模式為線性或非線性,皆可用 GRNN 來解決。

而 GRNN 時常用於函數逼近(Function Approximation),函數逼近的概念如下:假設 $f(x)$ 為 $x=[x_1 \ x_2 \ \dots \ x_d]$ 的函數,其函數逼近是以基底函數 $\varphi_i(x)$ 之線性組合來描述 $f(x)$ 。以 GRNN 而言,其基底函數 $\varphi_i(x)$ 是為 Radial Basis Function,即為公式(2):

$$\varphi_i(X) = e^{-\|w-p\|^2/b^2} \quad (2)$$

它的目的是將 summation Layer 的值做映射得到所需要的輸出。 w 為權重值向量, p 為輸入向量, $\|w-p\|$ 為計算權重值向量與輸入向量之間的歐幾里得距離(Euclidean Distance),之後乘上偏權值 b ,即成為轉移函數—高斯函數之輸入值。

此外,類神經的訓練就是在調整權重值 w ,使其變得更大或更小,來達成學習的目的。權重值越大代表連接的神經元更容易被激發,對類神經網路的影響也更大。GRNN 之類神經網路架構圖如圖 2 所示。

其中變數的定義: $\hat{Y}(X)$ 為 output,亦即所需要的結果; n 為訓練集資料數; x_i 為訓練集

中第 i 個輸入向量; y_i 為訓練集第 i 個輸出值; w 為權重值; σ 為最佳平滑參數;而 summation Layer 此部分是將每一個輸入與權重值相乘後做一加總的動作。圖 3 為高斯函數($\sigma=2$)示意圖 (蔡明倫,2002)【3】

2.4 分類與迴歸樹(Classification And Regression Tree ,CART)

決策樹擁有一個強大的優點,就是能夠將訓練資料中最好的變特徵先進行分割,所以可以在進行預測或分類時清楚地指出那一個特徵最重要,進而降低決策樹遇上太多類別或訓練樣本數目太小時,所產生的錯誤預估率。

分類與迴歸樹 (Classification And Regression Tree ,CART) 即為決策樹的一種,由 Brieman 等人於 1984 年提出,最初用來解決分類上的問題,主要分為分類樹與迴歸樹兩種分析方式,其差異在於處理目標變數 (Target

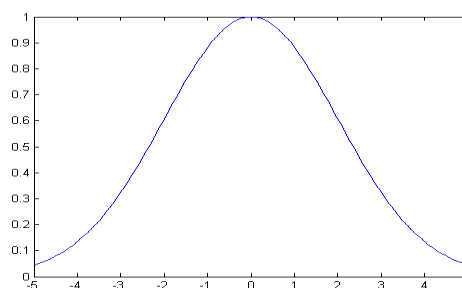


圖 3 高斯函數示意圖($\sigma=2$)

variable) 的不同而發展出來的，其中分類樹處理離散型資料，迴歸樹則處理數值與連續型資料，然後依照各自演算法分割條件，自動地將資料分類，並以樹狀圖的方式來表示。

CART 分類樹之分析程序如圖 4，首先將健檢資料劃隨機分兩群資料，一群做為訓練組樣本，另一群為測試組樣本，並以訓練組樣本藉著一個單一輸入變數函數，在每一個節點分隔資料，來建構一個節點數目最多的二分式決策樹，接下來透過測試樣本進行修剪分類樹的步驟，以產生許多的子樹，最後利用測試樣本作驗證來得到最佳的分類樹。

3.研究方法

本論文的脂肪肝檢測系統之系統流程圖如圖 5 所示。

在健檢資料部分，我們收集本論文所需要的相關資料後，會依據所需要的格式屬性去做調整、量化、正規化，所以資料處理方面可以分兩個部份來處理：

1. 遺漏值處理:當輸入資料庫的過程中有所遺漏或者是資料不完整時，本論文選擇直接刪除。
2. 資料正規化:將資料集轉換成程式所需要的格式。

資料探勘部分，因為它的目的包括分類 (Classification)、推估 (Estimation)、預測 (Prediction)、關聯分組 (Affinity Grouping)、

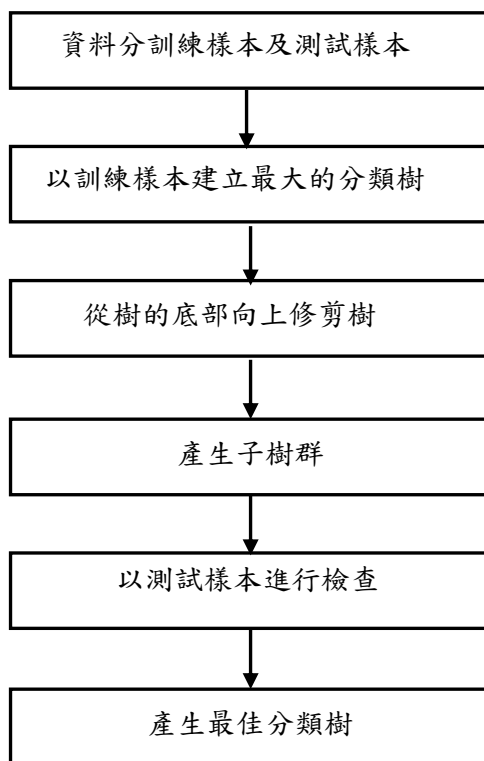


圖 4 分類樹分析流程圖

群集 (Clustering) 和描述 (Description)，所以會使用許多統計分析與 Modeling 的方法，到資料中尋找有用的特徵 (Patterns) 以及關連性 (Relationships)，而將資料中隱藏的資訊挖掘出來。本文則以分類 (Classification) 為主要功能。

Classification 是根據一些變數的數值做計

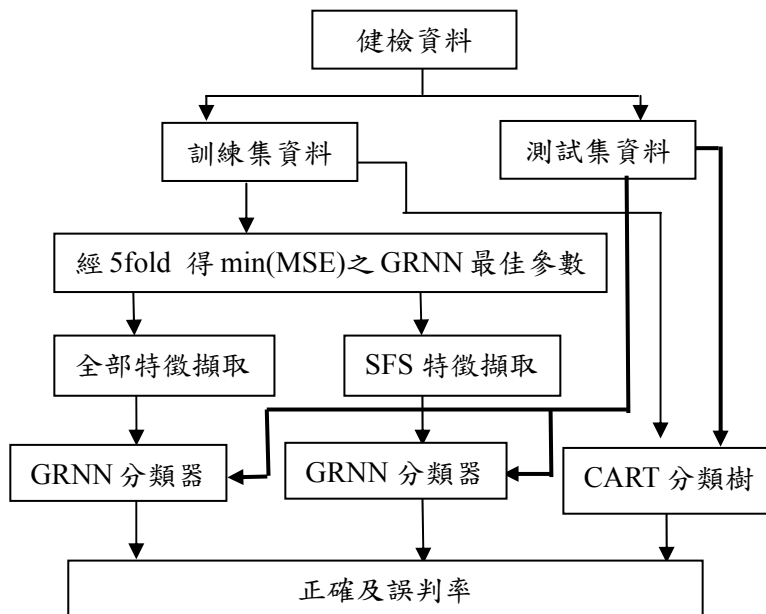


圖 5 脂肪肝檢測系統之系統流程圖

算，再依照結果作分類，然後擷取各個已分類資料的關鍵特徵作為分類參考，最後根據這些特徵對其他未經分類或是新的資料做分類預測。這些已分類資料可能是來自現有的資歷史性資料，或是將一個完整資料庫做部份取樣，再經由實際的運作來測試。

另外，資料探勘可分為兩種模式：一為假設檢定 (Hypothesis Testing)，是一種由上而下的方式，用以檢視我們的想法是否成立；另一為知識發掘，是一種由下而上的方式，從分析原始資料開始，找出所不知道的事實。本論文即為由下而上的資料探勘模式，目的在於分類脂肪與正常組織，所以論文中之特徵組合選擇是採用 SFS 來進行資料探勘，並搭配 GRNN 分類器及 CART 分類樹，產生不同分類模型來進行分析，以建立預測模式。

4. 實驗結果

本文實驗的主要工具是以 MATLAB R2006a 進程式撰寫，而 CART 的部份則是採用 SPSS Clementine7.2 進行設計。實驗用的數據是以中部某醫學中心所提供的 1133 筆健檢資料，以亂數分類方式分為訓練組 568 筆，測試組 565 筆。

4.1 實驗結果及分析

本文分類器之參數選擇是將訓練資料集，以 5-fold Cross Validation 量測在不同 σ 下之均方差 (MSE) 總和，其中最小均方差總和所對應之 σ ，即為後續特徵選擇時之分類器參數。表 1 為 GRNN 參數值數據資訊表，由表 1 可知本實驗 MSE=1.0786 時最小，其所對應之核心參數 $\sigma=2.5$ 。

特徵選擇方面，我們必須瞭解現實資料中，哪些是具有影響力的特徵，所以這些不同

表 1 GRNN 參數值數據資訊表

編號	5 群 MSE 加總	GRNN 參數 σ
1	1.0809	2.3
2	1.0789	2.4
3	1.0786	2.5
4	1.0796	2.6
5	1.0816	2.7

特徵變數間的關係需要有技巧性去操作，然後藉由不同特徵的選取方法 (Feature Selection Method) 來降低維度、刪除不必要的，最後選擇出對於脂肪肝最具代表性的特徵組合，使分類誤差降至最低。本實驗有三種分類模型，一是採用全部 43 個特徵建構於 GRNN 分類器，二是將 SFS 所挑選的特徵建構於 GRNN 分類器，三是利用利用 CART 分類樹模型的輸出結果。

其中 SFS 此方法是以均方差評量分類誤差，並將各維度中最小之均方差選擇出特徵進行排列，最後在所列的特徵中可選出最小均方差之特徵組合，而此特徵組合即為該次特徵選擇之結果。

圖 6 為測試集資料和訓練集資料的 SFS 分類器特徵選擇趨勢圖，由此圖可得知測試集資料大致為 U 型曲線，而訓練集資料是呈現遞減而趨於穩定的下降曲線，此兩曲線符合預期目標，且結果呈現 GRNN 經 SFS 特徵選擇採用 28 個特徵時，可產生最小的 MSE 值 0.209366，如表 2。

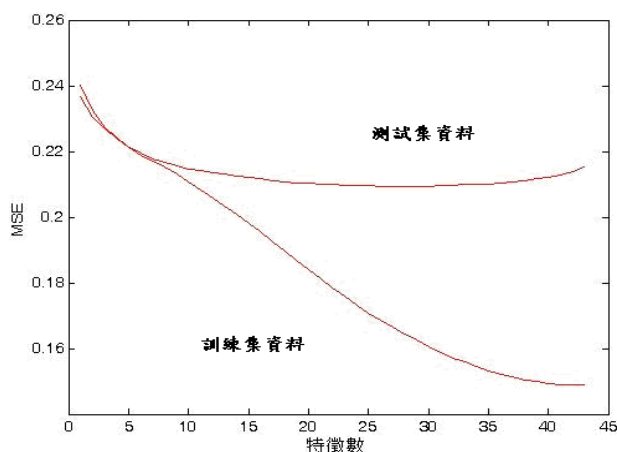


圖 6 SFS 分類器之特徵選擇趨勢圖

表 2 SFS 之 25-31 個特徵的 MSE 變化程度表

特徵個數	MSE 值	特徵編號
25	0.209500	15
26	0.209473	18
27	0.209404	30
28	0.209366	13
29	0.209374	25
30	0.209429	12
31	0.209477	24

表 3 為 SFS 特徵選擇建構於 GRNN 分類器上的特徵組合。由表 3 我們可以得知，GRNN 分類器經 SFS 特徵選擇，採用 28 個特徵時為最佳。

績效評估部分，本論文以表 4 所定義的受檢測區域與系統檢測結果之關係，產生公式 (3)(4)，然後以公式(3)產生的各分類器之正確率作為評斷。

$$\text{正確率(A\%)} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (3)$$

$$\text{誤判率(1-A\%)} = \frac{\text{FN} + \text{FP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}} \quad (4)$$

由於以一臨界值 Max (TP+TN) 作為判斷標準，所產生的正確率，才可與 CART 分類樹作相比較，因此，我們主要是將測試資料集分別投入上述三種分類模型中，然後採用相同 Max (TP+TN) 之狀況下，來產生本論文所需的正確率及誤判率。

表 5 呈現本實驗三種模型的正確率誤判率結果。由表 5 可知利用訓練集資料經 SFS 特徵選擇建構於 GRNN 分類器，再將測試集資料驗證此分類器，其正確率 0.7416 為最佳。因此就特徵擷取而言，經由 SFS 擷取特徵而建構的分類器，它的分類績效優於採用全部特徵而建構的分類器；就分類器而言，GRNN 分類器之輸入向量未經 SFS 特徵萃取，正確率劣於 CART 分類樹；反之，GRNN 分類器之輸入向量如經 SFS 特徵萃取，正確率則優於 CART 分類樹。整體而言，GRNN 分類器結合 SFS 特徵萃取之分類績效最佳。

表 3 SFS 特徵選擇建構於 GRNN 分類器上的特徵組合

	GRNN 分類器之 SFS 特徵選擇
特徵組合	9,4,39,36,8,32,2,7,38,3,23,11,43,28,26,41,40,10,14,42,34,22,29,31,15,18,30,13

表 4 受檢測區域與系統檢測結果之關係表

實際狀況 檢測結果	脂肪肝病患 (Disease Present)	正常人 (Disease Absent)
檢測為脂肪肝症 (Test Positive)	True Positive (TP)	False Positive (FP)
檢測為正常 (Test Negative)	False Negative (FN)	True Negative (TN)

表 5 三種分類模型正確率誤判率

	全部特徵+GRNN 分類器	SFS 特徵選擇+GRNN 分類器	CART 分類樹
特徵數	43 個	28 個	17 個
正確率	0.6867	0.7416	0.6938
誤判率	0.3133	0.2584	0.3062

5. 結論

本論文在特徵選擇方面，主要以 SFS 特徵選擇來執行特徵組合，而 SFS 此方法是以均方差評量分類誤差，並將各維度中最小之均方差選擇出特徵進行排列，因此在所列出的特徵中可選出最小均方差之特徵組合，接著使用此特徵組合建構於 GRNN 分類器。

實驗部份我們希望以 GRNN 分類器及 CART 分類樹，針對中部某醫學中心 2007 年度的相關健檢資料，經特徵擷取，在測試 43 組特徵組合後之最佳組合條件下，以正確率及誤判率來作其績效評估。所以我們分別針對全部特徵於 GRNN 分類器、經 SFS 特徵選擇於 GRNN 分類器、CART 分類樹等三種分類模型，使用測試集資料進行績效評估，其結果呈現經 SFS 特徵選擇建構於 GRNN 分類器，此一方式產生最高的正確率 0.7416。

未來研究發展，於資料探勘特徵選擇方面，實驗可測試以逐次反饋式搜尋法 (Sequential Backward Selection, SBS) 執行特徵選擇，或可使用逐次前饋式搜尋法結合逐次反饋式搜尋法成為 Sequential Floating Search Method (SFSM)以執行特徵選擇；分類器方面

往後可測試使用 GRNN 分類器進行脂肪肝之嚴重程度分類等級。

參考文獻

- [1] 賴世偉、廖光福、黃金財，”健檢民眾非酒精性脂肪肝盛行率之分析”，*輔仁醫學期刊*，第 3 卷，第 2 期，pp.75-80，2005。
- [2] 廖運範、陳東榮，”台灣肝臟病系列，四十三、脂肪肝”，*當代醫學*，第十九卷，第七期，pp.568-572，1992。
- [3] 蔡明倫，二微點狀影像資訊之強化、特徵萃取及辨識—以 X 光乳房微鈣化檢測為案例，未出版之碩士論文，私立大葉大學工業工程學系，彰化縣，2002。
- [4] 姚志成，運用資料探勘技術建構脂肪肝預測模式，未出版之碩士論文，私立中原大學資訊管理研究所，中壢市，2005。
- [5] 楊敦翔，以類神經網路與特徵選取技巧處理空氣能見度預測問題之研究，未出版之碩士論文，國立中山大學機械與機電工程學系，高雄市，2004。
- [6] 王元宏，以心電圖訊號檢測睡眠呼吸中止症的類神經網路，未出版之碩士論文，國立中山大學機械與機電工程學系，高雄市，

2004。

- [7] Specht, Donald F., “A General Regression Neural Network,” *IEEE transactions on Neural Networks*, Vol. 2, No. 6, pp.568-576, 1991.
- [8] Tanaka, S., Togashi, K., Rankinen, T., Perusse, L., Leon, A.S. and Rao, D.C. , “ Is adiposity at normal body weight relevant for cardiovascular disease risk? ”, *International Journal of Obesity*, Vol. 26, pp.176- 183 , 2002.
- [9] Yu, S. and Guan, L., “A CAD System for the Automatic Detection of Clustered Microcalcifications in Digitized Mammogram Films ,” *IEEE Trans. Med. Imag.*, Vol. 19, No. 2, pp.115-126, 2000.
- [10] Demsar J., Zupan B., Aoki N., “Feature Mining And Predictive Model Construction From Severe Trauma Patient’S Data”, *Journal Of Medical Informatic*, Vol. 63, pp.41-50,2001.
- [11] Ramirez, J.C.G., Cook, D.J., Peterson, L.L., Peterson, D.M., “Temporal Pattern Discovery In Course-Of-Disease Data”, *IEEE Engineering In Medicine And Biology Magazine*, Vol. 19, Issue. 4, pp. 63 –71, 2000.