

應用類神經網路於心血管疾病分類與診斷

學生：陳振亨

指導教授：胡坤德

林水順

東海大學工業工程與經營資訊研究所

摘要

隨著人們生活物質條件的不斷提高，社會文明及精神文明不斷進步，為了講求效率以致生活步調變快，飲食日漸西化，罹患心血管疾病的年齡也隨之逐漸下降。在台灣十大死亡原因中與心血管疾病相關疾病死亡率總和已達20.04%。所以尋找一個適合且準確的方式用以診斷心血管疾病是相當值得研究的一個課題。

為了讓國人能夠自行依據健康檢查報告判斷自己的健康狀況。本研究目的在於探討使用類神經網路系統應用於個人健康管理上，達到管理個人健康狀況，進而提升醫療品質與降低社會成本。

透過分析健診資料與生活習慣項目，運用類神經網路預測可能罹患的心血管疾病並提出適當建議。採用專家知識與類神經網路建構一套心血管病徵與診斷系統。將經由專家與共同資訊特徵選擇法(MIFS)篩選後的特徵，利用倒傳遞網路(BPN)讀取專家知識庫資料來分析預測並且提供有關心血管疾病患者日常生活飲食相關知識。

因此本研究希望達到下列三個目的：

1. 找出健診特徵與生活習慣對於心血管疾病的關連性。
2. 建構以專家知識與類神經網路為基礎的參考模式，並分析模式行為。
3. 設計心血管疾病分類與診斷系統雛型，並驗證其可行性。

關鍵字詞：類神經網路、倒傳遞網路

Applying neural network in classifying and diagnosing cardiovascular disease

Student : Zhen-Heng Chen

Advisor : Prof. Kun-Te Hu

Prof. Shui-Shun Lin

Department of Industrial Engineering and Enterprise Information
Tunghai University

ABSTRACT

With the constant improvement of people's material life conditions, social civilization and spiritual civilization are making progress continuously, however in order to have great efficiency, it already resulted in the life step becomes fast, diet occidentalize day by day, and the people age which suffers from cardiovascular diseases is declining. In reason of ten major deaths in Taiwan, the total death rate relevant to cardiovascular disease has already been up to 20.04%. Thereupon, looking for a proper and accurate way to diagnose the cardiovascular disease is a subject that is quite worth studying.

In order to let people judge their own health condition according to the health examination report by itself. This research purpose lies in the probe into the using type neural network system to apply to personal health management, to reach the personal health status management, and then improve medical quality and reduce the social cost.

Studying through the analysis of health examination and the living habit event, we can utilize the type neural network to predict the probability of suffering cardiovascular disease and then propose the proper suggestion. Adopting expert's knowledge and type neural networks to construct a set of cardiovascular symptoms of diseases and the diagnose system. The characteristics sorted out via the expert and MIFS, to utilize BPN to read expert's knowledge base materials for analysis prediction and provide the relevant knowledge of daily life diet for the people suffering cardiovascular diseases. The research objectives are three fold:

1. To find out the characteristics of the physical examination the habits and habits related to cardiovascular disease.
2. To construct a reference model based on expert system and neural network, so as to analysis the behavior.
3. To contract a cardiovascular disease symptom classification and diagnose system, and testify its feasibility.

Keywords : Neural Network 、 Back-propagation Network

致謝

在研究所兩年的生活中，是個具有酸甜苦辣的一段日子。回想起這段期間，有著許多的感觸。雖然一路走來很辛苦，不過也讓我成長不少。在這兩年之中的生活裡，首先要感謝的是我的指導老師胡坤德教授，他在課業上用心的指導，在人生的規劃上也給我許多的指引，讓我受益良多，在論文指導上提供寶貴意見與認真指導，在此，我由衷感謝胡老師的一切幫助與教悔。

在研二論文寫作的過程中，很感謝胡坤德及林水順兩位老師。兩位老師在指導論文過程中，讓每當我發現到有困難時都能夠順利的迎刃而解，還有心情的調適上讓我被感溫馨。在口試期間，感謝吳信宏博士及王文清老師特別撥空審閱論文，提供寶貴意見與指導，使的論文內容更為完備，在此感謝兩位老師的指正與幫忙。

另一方面，這兩年的研究生活中，特別感謝父母親及家人的關心與照顧，讓我能很順利的完成學業。此外，還要感謝同屆的同學兆廷的幫助，還有學弟哲鋒、昕煒、紹維的陪伴與支持，以及同研究室的諸位同學與學弟妹的鼓勵，讓我這兩年的生活更加充實與開心，謝謝你們！

最後還是特別要感謝我的父母及家人的關懷與照顧使我能在無憂無慮的環境下，順利完成學業。我想這份成果是屬於我們的。

陳振亨 謹誌於

東海大學工業工程與經營資訊研究所

中華民國九十四年七月

目錄

第一章 緒論	1
1.1 研究背景	1
1.2 研究動機	2
1.3 研究目的	3
1.4 研究範圍與限制	3
1.5 研究流程	4
1.6 論文架構	5
第二章 文獻探討	6
2.1 類神經網路	6
2.1.1 類神經網路的基本架構	7
2.2 類神經網路在醫藥相關研究與應用	11
2.3 心血管疾病	14
2.3.1 常見的心血管疾病	15
2.4 資料採礦 (Data Mining)	16
2.4.1 資料採礦之醫療相關研究與應用	17
2.5 倒傳遞網路的發展史	18
2.6 共同資訊特徵選擇法 (MIFS)	20
2.6.1 熵理論與共同資訊	22
第三章 研究方法與系統架構	24
3.1 資料範圍	24
3.2 研究工具	24
3.3 研究機制設計	24
3.3.1 資料搜集	28
3.3.2 資料集屬性挑選	28
3.3.3 原始資料前處理	30

3.3.4 建構類神經分類器及模式評估	32
3.4 倒傳遞網路運作過程	32
3.4.1 BPN 網路參數設定	37
3.4.2 BPN 神經網路學習與測試	40
3.4.3 預測準確度的評估標準	41
3.5 系統架構	42
3.5.1 系統主要元件	43
第四章 BPN 網路實作與系統試驗	45
4.1 病患特徵篩選	45
4.2 類神經分類器訓練與模擬	47
4.2.1 類神經分類器實做	47
4.3 系統設計與測試	55
4.3.1 系統設計	55
4.3.2 測試	58
第五章 結論與建議	60
5.1 研究成果	60
5.2 建議	61
參考文獻	62
附錄	65

圖目錄

圖 1.1 研究架構圖	4
圖 2.1 回饋式(feedback)類神經網路.....	10
圖 2.2 向前式(feedward)類神經網路.....	10
圖 2.3 從資料到決策的轉化過程.....	16
圖 2.4 倒傳遞網路架構	20
圖 3.1 BPN 網路研究機制設計	26
圖 3.2 BPN 網路之資料流程圖.....	27
圖 3.3 雙彎曲函數圖	30
圖 3.4 雙彎曲正切函數圖	31
圖 3.5 BPN 網路分類示意圖.....	32
圖 3.6 單階隱藏元網路架構	38
圖 3.7 系統架構圖	42
圖 3.8 使用者流程圖	43
圖 4.1 Resolution1.....	48
圖 4.2 Resolution2.....	49
圖 4.3 Resolution3.....	50
圖 4.4 分類器一	51
圖 4.5 分類器二.....	51
圖 4.6 分類器一誤差統計圖	53
圖 4.7 分類器二誤差統計圖	54
圖 4.8 輸入介面	56
圖 4.9 輸出介面	57
圖 4.10 測試資料系統介面.....	58
圖 4.11 測試結果介面圖.....	59

表目錄

表 1.1	民國 92 年十大死因依序排列.....	1
表 2.1	國內類神經網路在醫藥方面的研究與應用.....	13
表 2.2	國外類神經網路在醫藥方面的研究與應用.....	13
表 2.3	國內資料採礦技術應用於醫學發展史.....	17
表 2.4	國外資料採礦技術應用於醫學發展史.....	18
表 2.5	倒傳遞演算法發展史.....	19
表 3.1	原始資料項目如下.....	29
表 3.2	BPN 網路學習速率遞減之情形.....	39
表 4.1	MIFS 採用的變數項目.....	46
表 4.2	專家採用的變數項目.....	46
表 4.3	最終採用的變數項目.....	47

第一章 緒論

1.1 研究背景

根據衛生署2004年統計顯示，國人每年醫療費用高達5,200億，且每年以10%成長，表示國人往醫院看診之比例明顯偏高。隨著醫療品質的提升，人口老化的趨勢也愈來愈增加，文明病亦快速成長。根據行政院衛生署2004年計【3】，台灣人口老化比例已佔總人口數8.8%，文明病如惡性腫瘤、腦血管疾病、心臟疾病、糖尿病等已列為台灣十大死因，且有日益遞增現象，如表1.1所示。

表 1.1 民國 92 年十大死因依序排列

順位	死亡原因	死亡人數	百分比
1	惡性腫瘤	35,201	27.10
2	腦血管疾病	12,404	9.55
3	心臟疾病	11,785	9.07
4	糖尿病	10,013	7.71
5	事故傷害	8,191	6.31
6	慢性肝病及肝硬化	5,185	3.99
7	肺炎	5,099	3.93
8	腎炎、腎徵候群及腎性病變	4,306	3.32
9	自殺	3,195	2.46
10	高血壓性疾病	1,844	1.42

隨著人們生活物質條件的不斷提高，社會文明及精神文明不斷進步，

為了講求效率以致生活步調變快，飲食日漸西化，罹患心血管疾病的年齡也隨之逐漸下降。根據長庚醫院【13】2005年統計資料顯示，國人第一次心肌梗塞發作的平均年齡為五十二歲，男女比例為十比一，病人中因為冠狀動脈疾病引起的心肌梗塞的年紀最輕為二十二歲。

中央健保局【1】在2004年的統計資料中顯示全球心血管疾病死亡總數，近半是亞洲人。十大死因心血管疾病致死機率佔了三大。國人腦血管疾病、心臟疾病、高血壓性疾病相關疾病死亡率總和已達20.04%，僅次於惡性腫瘤。

根據行政院衛生署所發佈的「92年國人主要死因與主要癌症統計結果」資料顯示，台灣地區民眾平均每一百個人死亡中就有將近九個人是與心臟疾病有關，長期高居台灣地區主要死亡原因第三位，而心臟疾病之中，又以冠狀動脈心臟病(以下簡稱冠心病)，因為可能引起心肌梗塞或是心因性猝死，對健康的危害最大。因此，心血管疾病的診斷與預防是值得來研究的。

1.2 研究動機

雖然台灣已邁向開發國家之林，但對於健康資訊接收程度卻沒有提升。根據輝瑞大藥廠在2003年對台灣都會區民眾所進行的「心血管疾病和膽固醇認知調查」【24】顯示，不到一成的受訪者有意接受心血管疾病檢查，八成的人未曾聽過高、低密度脂蛋白膽固醇，是亞洲國家中排名最後。國人嚴重缺乏對於心血管疾病的認知。

衛生署【3】所調查統計的數據全國有一百四十五萬名心臟病患、六十五歲以上的老人則是五人就有一人是心臟病患者，顯見心臟病對國人健康的威脅日益加重。同時調查中也得知，罹患心臟病者亦是高血壓、糖尿病、高血脂等三高疾病的高危險群。通常病發時才會尋求醫療協助。未能及時防止，耗費了相當多的醫療資源與社會成本。

在台灣十大死亡原因中與心血管疾病相關的，即包括了三項：腦血管疾病、心臟疾病與高血壓性疾病。心血管疾病是目前常見導致死亡和喪失活動能力的最常見原因，所以尋找一個適合的方式用以輔助診斷心血管疾病是相當值得研究的一個課題。

1.3 研究目的

資料採礦在於醫療方面的應用多為臨床路徑分析、病因研究、疾病與用藥相關之研究、刪除浪費降低成本。使用者多為醫院管理階層或具相關醫學知識的醫師。一般資料採礦醫療資訊系統多為專業人士所設計，與一般國人的使用需求差異相當大，沒有具備專業知識難以操作與判讀。綜合研究背景及研究動機，為了讓國人能夠自行依據健康檢查報告判斷自己的健康狀況。本研究目的在於探討使用類神經網路系統應用於個人健康管理上，達到管理個人健康狀況，進而提升醫療品質與降低社會成本。

因此本研究希望達到下列三個目的：

1. 找出健診特徵與生活習慣對於心血管疾病的關連性。
2. 建構以專家知識與類神經網路為基礎的參考模式，並分析模式行為。
3. 設計心血管疾病分類與診斷系統雛型，並驗證其可行性。

1.4 研究範圍與限制

以一般大眾之個人健康管理系統架構為主要觀點。影響健康的因素有很多，包含生活習慣、遺傳病、運動、居家環境、用藥習慣、工作壓力、生理因素、心理因素等許多方面。由於時間、人力、成本等因素的限制，無法對全部因素加以探討，僅對健康檢查報告與生活習慣探討。

1.5 研究流程



圖 1.1 研究架構圖

1.6 論文架構

本研究共分為五章，依序為緒論、文獻探討、系統分析與設計、系統實作、結論與建議，各章主要內容分述如下：

第一章：緒論。本章主要介紹研究動機、背景、目的、限制、流程，並說明本論文之整體架構。

第二章：文獻探討。介紹類神經網路於醫藥的發展與應用、心血管疾病研究、常見的心血管疾病、介紹資料採礦在醫學方面的發展與應用、以及類神經網路、倒傳遞演、共同資訊特徵選擇法之相關文獻探討。

第三章：研究方法與系統架構。病患特徵的項目篩選、資料轉換、BPN網路的設定、系統架構、系統分析、系統設計、系統實做。

第四章：BPN網路實做與系統試驗。依照前章BPN網路所架構的模型進行分析。心血管疾病預防系統之測試。

第五章：結論與建議。總結系統所達成之目標，並說明系統未來可進一步之研究方向。

第二章 文獻探討

2.1 類神經網路

類神經網路最早在1890年提出，直到1943年，才在McCulloch 和Pitt 提出運算元數學模式，1957年由俄國數學家Kolmogorav 證明出理論，而後因學習能力的受限，使得類神經網路受限而沈寂，直到1980年代，類神經網路才有突破性的發展，又開始為人們所重視【27】。

類神經網路是「一種基於腦與神經系統研究所啟發的資訊處理技術」。它可以利用一組範例（即系統輸入與輸出所組成的資料）建立系統模型，這樣的系統模型便可用於推估、預測、決策、診斷。常見的迴歸分析統計技術正是利用一組範例（即數據）建立系統模型（即迴歸公式），因此類神經網路也可看成是一種特殊形式的統計技術。由於類神經網路已被應用於領域，並且有著許多成功的例子，再加上近來電腦技術的進步，大幅節省運算所需時間，因此類神經網路的應用也越趨廣泛。類神經網路具有以下之特性【27】：

1. 容錯能力：

傳統的系統當輸入資料殘缺矛盾或雜訊過多時，常無法獲得正確的結果，但類神經網路則沒有這些缺點，它因運算元及連結鍵組成的方式，特別擅長處理錯誤、多餘、矛盾、重複、殘缺的資料，也就是具模糊（Fuzzy Reasoning）推理能力。

2. 聯想性記憶：

類神經網路可以根據過去的經驗，以自我學習的方式判斷找出相當接近真實的結果。藉著自我學習，解決使用者在資料的描述及處理上所遇到的一些缺乏答案的問題。

3. 平行分散式處理方式：

傳統的數位電腦式採用循序的處理方式，依其程式內容依序運作，所以不管記憶體之容量多大，一次也只能運作一部份，而類神經網路在

理論與實際上，都是採用平行分散性的方式來處理資訊，能夠迅速、簡單地處理資訊。

4. 具有處理非線性訊號的特質：

類神經網路可從輸入及輸出資料中相互對應學習，在線性系統中為單一輸入和輸出的變化，且輸入值為單一的影響，但在非線性系統中各輸入值為相互關連，輸入值彼此會相互影響。在真實世界中許多問題往往是非線性的，類神經網路提供了處理實際複雜非線性系統的功能。

5. 儲存知識的特殊方式：

傳統的電腦是將所獲得的知識儲存在記憶體中特殊的位置，而類神經網路則是將學習所獲得的知識儲存在互相連接的運算元間及連接鍵中；換句話說，它的知識是存於其結構中，而非特定位置的記憶體中。

6. 即時性分析：

類神經網路在學習過程中，學習的時間通常很長，以便獲得較佳的訓練結果，但在回憶過程時卻很迅速，故網路一經建立後，適用於即時性分析。

2.1.1 類神經網路的基本架構

類神經網路是由許多個運算元所構成。一個類神經網路之架構是由運算元之連結方式、轉換函數和學習法則來決定。類神經網路之架構說明如下【25】。

1. 運算元(Processing Element)：

亦可稱為處理單元，為組成類神經網路之最基本單元。運算元主要是接受輸入值後彙總將其輸出，運算元的作用可用三種函數來說明。

(1) 集成函數(Summation Function)

集成函數是用以將從其他處理單元之輸出透過網路連結傳來的訊息加以綜合。如加權乘積和、歐式距離等.....。

(2) 作用函數(Activity Function)

目的將經由集成函數傳來的輸入和該運算元目前的狀態加以彙總。作用函數之功能不明顯所以一般類神經網路大多採取集成函數之輸出。

(3) 轉換函數(Transfer Function)

轉換函數的目的在把作用函數之輸出值轉換成運算元之輸出值。

如：屬於線性函數的Perceptron 函數(或稱階梯函數Step Function)、Hopfield-Tank 函數、Signum 函數及非線性函數的雙彎曲函數、雙曲線正切函數等等.....都是在類神經網路上在使用的轉換函數。

2. 層(Layer)：

由數個功能相同的運算元組成，可區分為三種，分別是輸入層、隱藏層及輸出層。層本身也有三種功能，描述如下：

(1) 正規化輸出：

正規化輸出之目的在於將位於同一層中的處理單元的原始輸出值所組成的向量加以正規化，成為單位長度的向量後，再作為層的輸出。

(2) 競爭化輸出：

競爭化輸出的目的在於將位於同一層中之處理單元的原始輸出值所組成的向量中，選擇一個或數個較強值的處理單元，令其值為1，其餘值為0後，再作為此層的輸出。這些輸出值為一的處理單元稱為優勝單元(Winner)。

(3) 競爭化學習

競爭化學習將位於同一層中之處理單元的原始輸出值所組成的向量中，選擇一或數個優勝單元，而後網路只調整這些與優勝單元相連的下層網路連結。

3. 連接鍵(Connection)：

為類神經網路基本要素之一，其主要作用為連接層與層間的運算元，做為其傳遞訊號路徑之用。

4. 網路(Network)：

將層與層中之運算元利用連接鍵相互連接而成網路。根據處理單元連接的方式，可分為向前式(Feedward)和回饋式(Feedback)兩種架構。

圖2.1為回饋式(Feedback)類神經網路示意圖、圖2.2回饋式(Feedback)類神經網路示意圖。

向前式架構的運算元分層排列，形成輸入層、隱藏層和輸出層，每一層只接受前一層的輸出做為輸入；回饋式的輸出層之輸出會回饋到輸入層，或者層內各處理單元間有連結者，或者運算元不分層排列，只有一層。類神經網路中網路的運作模式可以分成學習過程與回想過程兩種【27】。

(1) 學習過程(Learning)：

不同的網路模式有不同的學習法則，在學習過程中，不斷的調整網路中的權數，若網路達到穩定狀態，學習過程就可以中斷。而學習的演算法又可分為三種。

a. 監督式學習(Supervised Learning)

在學習的過程中，每一個輸入之資料均有一個期望之輸出值對應，學習的目標在調整權數以降低網路輸出值與期望值之差距，由於需要不斷的學習、調整。一般而言，監督式學習在學習過程中需要花較時間才能得到一個較好的結果。

b. 非監督式學習(Unsupervised Learning)

在非監督式學習網路中，每一個輸出單元代表一個聚類結果，因此網路學習的目的在降低網路優勝單元的連結加權值所構成的向量與輸入向量間之差距。學習之品質和監督式學習一樣可以用能量函數來表示。利用此種學習方式的網路有自適應共振網路等。

c. 聯想式學習(Associative Learning)

此種學習方法是利用運算元組成的向量來表示模型(Pattern)，網路學習旨在使從一開始所組成的初始樣式，經過聯想的過程得到最終樣式。此種學習方式又可分為自聯想學習與異聯想學習兩種，例如：霍普菲爾網路與雙向聯想記憶網路等。

d. 回想過程(Recall)

經過學習過程的學習之後，類神經網路會將學習之結果儲存在網路的连接權數中，因此我們將資料輸入網路之後，經過一些數學運算即可獲得輸出結果，此過程稱之為回想過程。

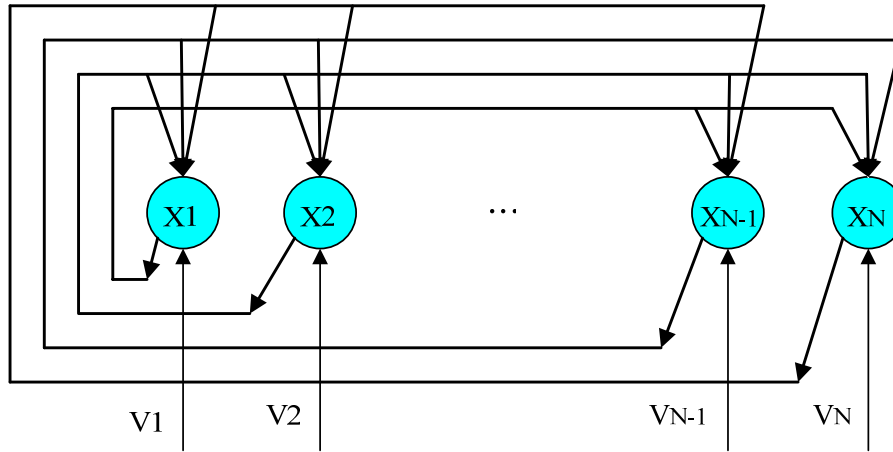


圖 2.1 回饋式(Feedback)類神經網路

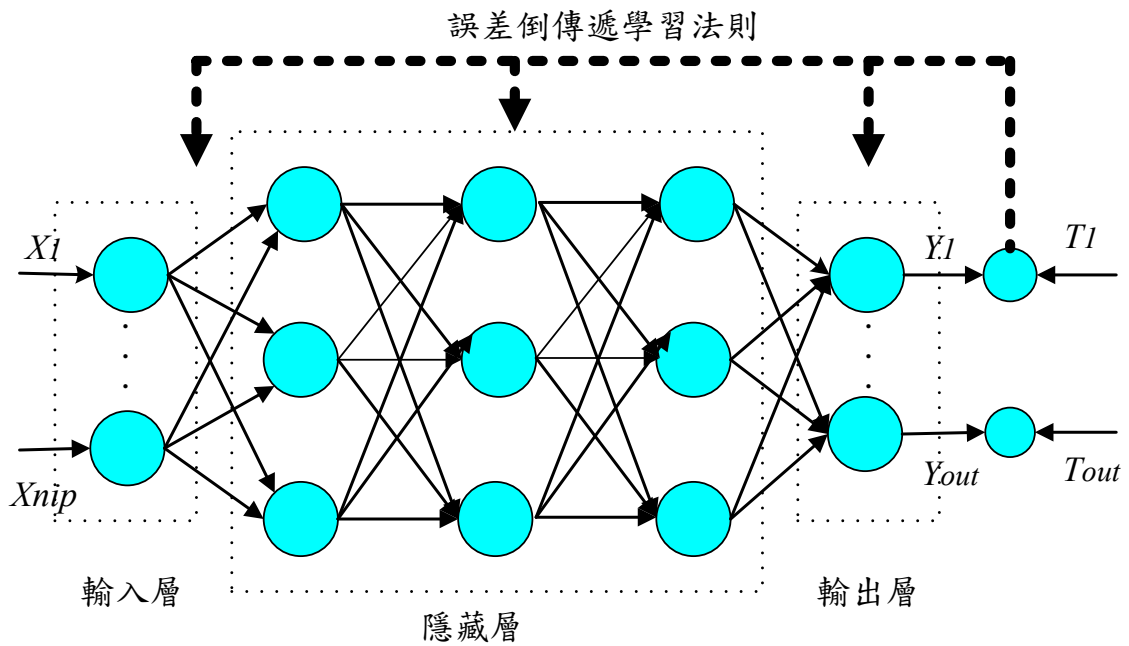


圖 2.2 向前式(Feedward)類神經網路

2.2 類神經網路在醫藥相關研究與應用

審查制度一直為各級醫療院所抱怨，而另一方面醫療費用仍不斷成長；顯見目前審查制度無法達到有效控制醫療費用，使其合理申報之功能。如何以現有的申報資料去了解各醫療院所之醫療品質是一個重要課題。依目前審查辦法，各抽審案件均交由各專科醫師審查，雖有分科會議討論爭議案件，但並無共同審查指標的建立。中央健保局在 1998 年【1】所採用之技術，乃電腦人工智慧一類神經網路之技術，其解決問題之過程係透過對專家意見及核減資料之特徵的瞭解而歸納出共同審查指標。

尋求申報資料與醫療院所行為之內含關聯性，因門診與住院申報資料不斷擴增，審查醫師的審查負荷也逐漸加大，審查醫師必需採取客觀審查指標才能了解申報資料的整體形態。中央健保局在 1999 年【1】提出以疾病別為基礎用類神經網路電腦篩選其醫療行為指標，即是以微觀方式選取異常之申報個案，如能直接以電腦篩選方式針對某些疾病建立其申報常模，由專科醫師了解該特定疾病之申報特性，則審查醫師可以更客觀的態度去建立審查指標。

孫漢屏在 2001 年【14】以類神經網路中的倒傳遞網路演算法，建構一類神經網路掛號專家系統，其系統可讓病患在使用網路預約掛號作業中，依疾病症狀建議至適當的科別就診，以提升病患的醫療滿意度與減少醫療資源的浪費。

採用倒傳遞類神經網路(Backpropagation Neural Network)建立燒傷患者住院日預測模型。實驗過程中將住院早期(急性期)階段分為「住院初期」與「療程初期」，住院後期階段則分為「出院準備期」，運用十摺交叉驗證法(10-fold Cross Validation)進行類神經網路網路學習。藉由分析住院早期的預測結果，以了解不同住院階段所採用之不同資料變數，對預測準確率的影響。同時亦參考住院後期之預測結果，以進一步驗證住院早期預測結果之參考價值與可

行性，希望能在患者住院早期就有較佳之預測結果，藉此達到具有時效性之住院日管理機制【17】。

雷賀君在 2003 年【21】提出為了解決資料的不確定性造成的困難以及維持資料分類規則的準確性，結合了粗略集合理論(Rough Set)與基因演算法 (Genetic Algorithm)作為資料探勘分類及預測的工具，有效的處理資料的不確定性問題，建立資料間之相關性以作為決策之依據，並結合類神經網路(Neural Network)作為資料分類規則的學習機制。其研究將以造成前十字韌帶傷害的因素作為討論之對象，運用資料探勘技術，從大量資料中找出前十字韌帶傷害與其他膝關節傷害間的關聯性，以及造成前十字韌帶傷害的關鍵因素。

朱佳雯在 2003 年【5】提出一個以人工智慧的混合型案例式推理所建構的心電圖自動診斷系統。在案例索引的部份，以類神經網路在監督式學習模式中的倒傳遞演算法進行案例分類，加快了搜尋相似案例的速度。在案例相似度的計算上，使用前述類神經網路在訓練時所得到的權重值，來決定每個特徵值的重要性。然後利用有權重的相似分數來計算輸入案例與存放於案例庫中既有案例的相似性。

唐壽生在 2004 年【16】運用資料探勘技術分別以迴歸分析與類神經網路二種方法，建立肺結核病患完治預測模式。在相同的條件下，發現類神經網路模式之預測能力普遍較佳，其平均誤差均方(MSE)值也比迴歸分析之 MSE 值低。如果未來疾病管制局能夠提供全國影響肺結核病患完治的影響因素的資料，再藉由資訊科技的輔助，相信必能建立更為精準的肺結核完治預測模式，以提高實用性。

表 2.1 為本研究整理近年國內類神經網路在醫藥方面的研究與應用，表 2.2 為本研究整理近年國外類神經網路在醫藥方面的研究與應用。

表 2.1 國內類神經網路在醫藥方面的研究與應用

學者或機構	研究或應用
中央健保局 (1998)	類神經網路之技術應用於醫療費用審查制度
中央健保局 (1999)	疾病別為基礎用類神經網路電腦篩選其醫療行為
孫漢屏(2001)	類神經網路為基礎之智慧型醫院網路掛號系統
袁繼銓(2002)	以類神經網路預測燒傷病患住院日之研究
朱佳雯(2003)	案例式推理與類神經網路在心電圖診斷之應用研究
雷賀君 (2003)	前十字韌帶傷害快速診斷系統-以粗略集合、基因演算法與倒傳遞網路為工具
唐壽生(2004)	資料探勘技術應用於肺結核病患完治的預測

表 2.2 國外類神經網路在醫藥方面的研究與應用

學者	研究或應用
Ichimura, Takum (2003)	A classification capability of Reflective Neural Networks in medical databases
Chang, Chuan-Yu (2004)	A Contextual-based Hopfield Neural Network for medical image edge detection
Leong, Jenn-Hwai 等 (2005)	A hybrid neural network system for pattern classification tasks with missing features
Guler, Inan (2005)	Improving medical diagnostic accuracy of ultrasound Doppler signals by combining neural network models

2.3 心血管疾病

對心臟血管疾病方面的研究中，較著名的流行病學世代研究為1950年代於美國開始的佛來明罕研究(Framingham Heart Study)【36】，經長期追蹤的結果發現，血脂質中的總膽固醇(Total Cholesterol, TC)、三酸甘油酯(Triglycerides, TG)與低密度脂蛋白膽固醇(Low-Density Lipoprotein Cholesterol, LDL-C)的濃度越高，冠狀動脈疾病的死亡率越高；而高密度脂蛋白膽固醇(High-Density Lipoprotein Cholesterol, HDL-C)濃度越高，冠狀動脈疾病的死亡率則越低【4】。

此外，生態研究中的Keys的七國研究(Seven Countries Study)Keys 在1984年【42】指出，血中總膽固醇值越高，冠狀動脈心臟病的死亡率越高，並且飲食中飽和脂肪佔熱量的百分比與冠狀動脈心臟病的死亡率呈很強的正相關(相關係數 $r=0.84$)【41】。以日本移民為研究對象的Ni-Hon-San Study【42】，其結論顯示當具有同一種族特質的研究族群，隨環境的變遷，以致飲食中的飽和脂肪佔熱量的百分比提高時，會使得血中總膽固醇值增加，而心血管疾病的發生率也隨之提高。以飲食進行調控血脂質的挪威奧斯陸心臟研究(The Oslo Heart Study)【18】，將1232位罹患高膽固醇血症的男性患者分成兩組，一組接受飲食教育與戒菸，另一組為對照組。經過四年的追蹤，結果顯示實驗組比對照組的血清總膽固醇濃度低13%、三酸甘油酯低20%、吸菸量減少45%，而罹患心肌梗塞的機會降低47%。除了上述這些著名的流行病學研究外，研究均指出，血脂質異(Dyslipoproteinemia)【28】與飲食是罹患心臟血管疾病相當重要的兩項危險因子【22】。

2.3.1 常見的心血管疾病

常見的心血管疾病包括了高血壓、冠心病、充血性心臟衰竭、中風等。將近五成的心血管疾病發作是突發的，因此平時勤加保養及定期健康檢查成為預防心血管疾病的最佳方式。在定期檢查方面，利用傳統健檢可測知血壓、三酸甘油脂、總膽固醇、低密度脂蛋白膽固醇、高密度脂蛋白膽固醇、心跳頻律等基礎指標【9】。

1. 高血壓

「高血壓」即是血管內的血液對血管壁的壓力超過了正常壓力值。高壓即收縮壓，又叫心縮壓，是當心臟收縮把血液打到血管所測得的血壓；低壓即舒張壓，又叫心舒壓，是心臟在舒張所得的壓力。一般高壓持續在 140mmHg 或以上，低壓持續在 90mmHg 或以上，就是患有高血壓病了。原因不明的高血壓，佔 90%，可能與環境有關，如鹽份攝取太多、肥胖等長期經年累月的結果。續發性高血壓，是如腎臟病等其他疾病引起。高血壓剛開始不會有症狀，有些人會感覺到頭痛、頭暈失眠、呼吸短促、頸部酸痛等症狀時【4】。

2. 冠心病

在冠狀動脈血管內，脂肪、膽固醇和鈣組成一種粥樣斑塊，粘附在血管的內壁，造成血管狹窄，使血液流通不暢，從而發生心肌缺血，產生冠心病的症狀。冠心病患者，在身體疲勞、精神緊張、緊急應激的情況下，因發生血管的收縮，加上原有的血管狹窄，更會影響冠狀動脈內的血液流通，從而引起胸痛，如心前區絞窄樣疼痛，甚至放射到左肩、左上肢前內側，俗稱心缺血，即時的休息，胸痛症狀可以減輕。但是當某一條血管發生完全阻塞時，即時的休息都不能緩解胸痛，則就發生了心肌壞死，需要即時搶救【4】。

3. 充血性心臟衰竭

心臟是一種「幫浦」靠反覆收縮及舒張，將血液送到肺臟及全身各器官組織。以應付體內代謝需要。當心臟不能正常執行這種功能，稱為心臟衰竭。心臟衰竭原因如下【4】：

(1)低排量心臟衰竭：因心臟本身的毛病，如冠狀動脈疾病、高血壓病、心肌梗塞、瓣膜性疾病、血色疾病造成心臟衰竭，其心輸出量 bvg 通常是降低稱為「低排量心臟衰竭」。

(2)高排量心臟衰竭：如甲狀腺機能亢進，貧血，激烈運動，代謝需求增加稱為「高排量心臟衰竭」。

4.中風

中風指部份腦部細胞因腦部動脈「閉塞」或「爆裂」，腦細胞因得不到營養及氧氣而死去【4】。

2.4 資料採礦 (Data Mining)

Data Mining 又稱為資料採礦、資料挖掘或資料探勘。資料採礦係指利用分類、關連性、序列分析、群集分析、機器自我學習及其它統計方法，自龐大的資料庫中找出隱藏、未知的，但卻對企業經營十分有用的資訊。有些企業的歷史資料都以百萬、千萬計，要分析起來相當困難，而利用資料探勘工具，可從龐大的資訊中萃取出有用的資訊，以客觀的統計方法，快速且正確地找出企業需求的經營資訊，得到正確的銷售模式、客戶關係、行銷策略等，以利企業掌握正確經營動態，增加利潤並減少支出【7】。

「資料探勘」為一種「資料轉化」的過程，是由沒有組織的數字與文字集結的資料 (Data Set)，轉化為資訊 (Information)，再轉為知識(Knowledge)，最後才產生決策 (Decision) ，如圖2.3所示。

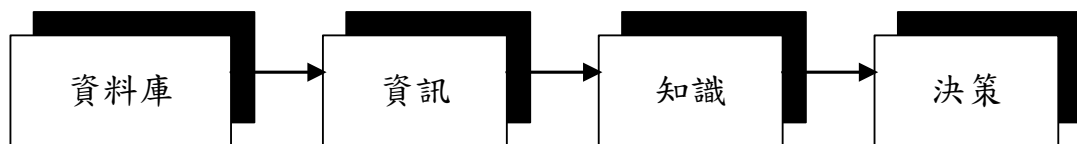


圖 2.3 從資料到決策的轉化過程

2.4.1 資料採礦之醫療相關研究與應用

黃仁貴在2002年指出【11】大醫院每日累積了大量門診、住院、檢驗、檢查等重要病患資料，但是資訊多樣性與無結構化之特性，使得醫院決策單位之護理人員或醫生，無法有效利用現存的資料轉換成有效的資訊，甚至使決策行為產生重大的損失或錯誤；如果能讓病患或護理人員透過電腦，及時且精確的提供個人化病歷資料，透過知識之表現與規則之建立，以資料探勘和資料倉儲之整合技術，解決大量病患歷年資料儲存之空間問題及效率問題，如此可建立個人化專屬健康資訊與病歷系統。

將醫師開藥的處方簽建立成資料庫，再利用資料採礦來找出資源浪費的原因。除此之外，在其它研究中還有依據長期病的病徵來找尋各病徵之間的關係，幫助醫師進行疾病的診斷。甚至還有一個較為先進的應用，就是基因序列的解析【10】。

蔣肇慶在1999年【26】提出使用關連性規則，在一個病患的醫令資料倉儲裡，探究有相同疾病的分類碼ICD-CM (International Classification of Diseases, Clinical Modification)的不同病患，找尋相關聯的醫令項目，作為各個病例的基本醫令組，成為適合我國本土化的AP-DRGs(The All Patient DRGs)。歸納整理「資料採礦技術」運用於醫療產業之研究如表2.3、表2.4所示。

表 2.3 國內資料採礦技術應用於醫學發展史

學者	題目
蔣肇慶等 (1999)	資料開採在醫療資訊的研究
湯玲郎等 (2000)	資料萃取法在健保費用稽核之研究
黃勝崇 (2001)	資料探勘應用於醫療院所輔助病患看診指引之研究
吳國禎等 (2002)	資料探索方法在醫學資料庫之評估
李淑芬 (2002)	臨床路徑之建立機制—應用資料採礦技術

黃仁貴 (2002)	以叢集分析技術探討病患就診屬性與看診時間之關係
楊超然 (2004)	利用文件及影像檢索建立胃癌診斷與治療的案例式推理

表 2.4 國外資料採礦技術應用於醫學發展史

學者	研究或應用
Megaputer(2002)	Medical Fraud Detection Through Data Mining
Shaout (2003)	A fuzzy data mining system for medical records
Aslandogan (2004)	Evidence combination in medical data mining
Lehmann (2005)	Automatic categorization of medical images for content-based retrieval and data mining

2.5 倒傳遞網路的發展史

倒傳遞網路 (Back-propagation Network, BPN) 模式是目前類神經網路學習模式中最具代表性，應用最普通的模式。1957 年 Rosenblatt 提出感知機模式以來，研究者總覺得這種模式缺乏隱藏層的學習演算法，其學習能力受到很大的限制。人工智慧的創始人之一的 Minsky 在 1969 年更以專書討論這各問題，並指出無隱藏層的感知機模式無法解決互斥或 (eXclusiveOR, 簡稱 XOR) 問題，但無法解決隱藏層問題【27】。

Werbos 於 1974 年在其博士論文中提出了隱藏層學習演算法。1985 年 Parker 再次提出倒傳遞網路，同年 Rumelhart, Hinton and Willians 發表一篇倒傳遞網路文章才使其廣為人知，這篇文章可能也是類神經網路研究史上迄今被引用頻率最高的一篇論文。倒傳遞類神經網路基本原理是利用最佳陡坡降法 (The Gradient Steepest Descent method) 的觀念，將誤差函數予以最小化。

倒傳遞類神經網路屬於監督式學習網路，因而適合診斷、預測等應用。倒傳遞網路的網路架構如圖2.4 所示，包含了輸入層、隱藏層及輸出層，而隱藏層可以不只一層。同一層中的神經元彼此並不相連，而不同層間的神經元則彼此相連，且信號的流向是由輸入層向輸出層單向傳播。倒傳遞網路中的神經元，其最常用的非線性轉換函數為雙彎曲函數（Sigmoid Function），其中

$$f(x) = \frac{1}{1+e^{-x}}$$

這種函數有一種特性，及當 x 趨近於正負無窮大時， $f(x)$ 趨近於0或1，而 $f(x)$ 的值則介於(0,1)之間。表2.5為本研究所整理倒傳遞演算法發展史。

表 2.5 倒傳遞演算法發展史

年代	研究史上的大事
1957	Rosenblatt 提出認知機（Perceptron）模式，是最早的類神經網路模式
1969	Minsky 與Papert 出版「認知機」一書。
1974	Werbos 在博士論文中提出了隱藏層學習演算法。
1985	Parker 再次提出倒傳遞網路
1985	Rumelhart 等人提出倒傳遞網路（BPN）與通用差距法則（Generalized Delta Rule）

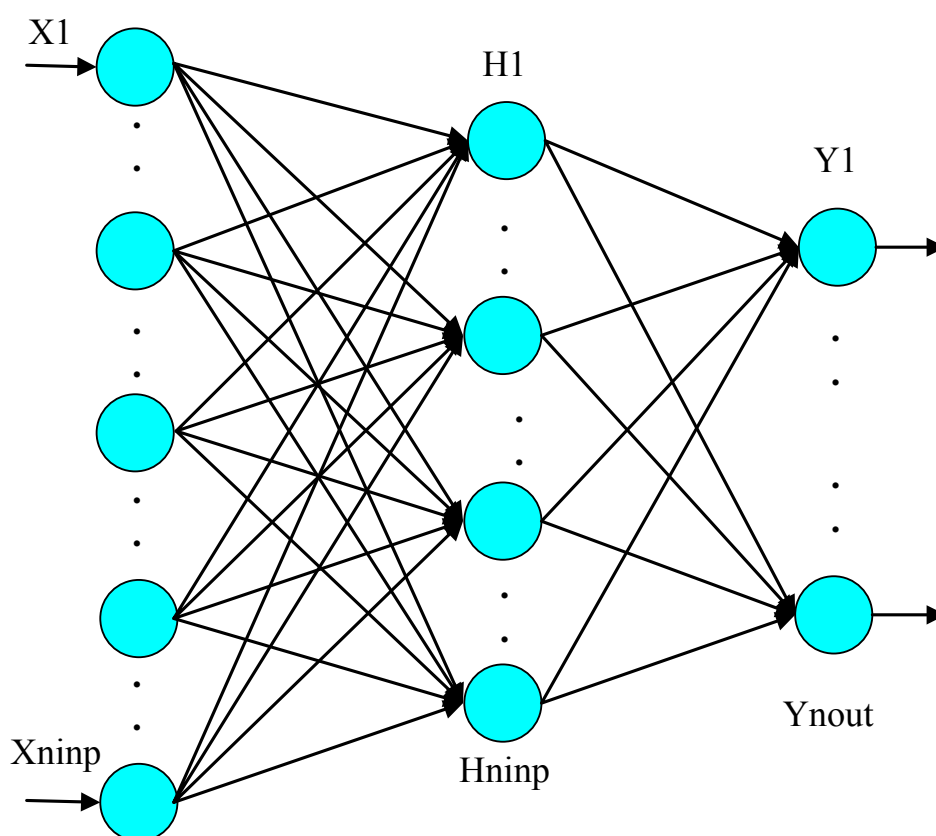


圖 2.4 倒傳遞網路架構

2.6 共同資訊特徵選擇法 (MIFS)

建立適當的類神經網路需決定除適當的網路架構、演算程序及參數設定之外，另一個重要課題為選取適當的特徵項目。特徵項目的選取對結果有深遠的影響。若選取的特徵項目不足或不當，便無法反映真實系統運作；若特徵項目太多，則造成網路學習的負擔。因此在進行網路學習前，應先擷取適當的特徵項目，其足以描述問題特徵。若問題的特性為固定形式，可由領域專家依據經驗篩選足以描述問題的特徵項目；若問題的特性會隨環境而變動，則需以科學性方法來找到適當的特徵項目【25】。以下將介紹共同資訊特徵選擇法(Mutual Information based Feature Selection, MIFS)。

Battiti根據一般的資訊理論，提出共同資訊特徵選擇法【30】。MIFS 演算程序以資訊理論中的熵理論與共同資訊做為計算的基準，評估原始資料集中特徵項目間相對重要程度。 F 表原始特徵項目之集合； S 表已選定之特徵項

目集合； f_i 表 F 集合中第 i 個特徵項目； $I(C:f_i)$ 表輸出變數 C 與第 i 個特徵項目之共同資訊； $I(f_i:s_j)$ 表 F 集合中第 i 個輸入變數與 S 集合中第 j 個輸入變數間之共同資訊。MIFS 之演算程序如下：

1. 初始設定： F 為原始資料中 N 個特徵項目之集合； S 為空集合。
2. 計算 F 集合中每一特徵項目之 $I(C:f_i)$ 值。
3. 選取特徵項目 f_i ，其 $I(C:f_i)$ 為最大值，令 $F \leftarrow F \setminus \{f_i\} ; S \leftarrow \{f_i\}$ 。
4. 重複以下步驟，直至 S 集合中含有 k 個選定的特徵項目：
 - (1) 計算任一組合 $(f_i, s_j) ; f_i \in F ; s_j \in S$ 。
 - (2) 找出令 $I(C:f_i) - \beta \sum I(f_i, s_j)$ 最大化之特徵項目 f_i ，並令 $F \leftarrow F \setminus \{f_i\} ; S \leftarrow \{f_i\}$ 。
5. 結束演算程序。

其參數 β 值愈大，表愈重視已選定特徵項目與未選定特徵項目間資訊重複性。Battiti 於研究中指出， β 值在 0.5 至 1.0 的範圍內可有不錯的分類表現。以下為輸出變數與輸入變數的 MI 值 $I(C;F)$ ，已選定輸入變數與未選定輸入變數的 MI 值 $I(X;Y)$ 的計算方式【30】：

$$I(C;F)=I(F;C)=\sum_c \sum_f P(c,f) \log \frac{P(c,f)}{P(c)P(f)} \quad (\text{公式2.1})$$

$P(c), P(f), P(x)$ 與 $P(y)$ 代表某一變數在樣本中的機率值， $P(c,f)$ 與 $P(x,y)$ 為二變數於樣本中的聯合機率值 $I(X;Y)$ 可整理如【40】：

$$I(X;Y)=I(Y;X)=\sum_x \sum_y P(x)P(y|x) \frac{P(x,y)}{P(x)P(y)} \quad (\text{公式2.2})$$

$P(y|x)$ 表觀察到變數 X 之後，變數 Y 發生的條件機率。經過 MIFS 演算程序的前處理，可減少需要的網路資料輸入量，大幅降低網路運算的負荷。其運作效果較其他特徵選擇計量方法（如主成份分析法 (Principal Component Analysis, PCA) 更好【30】。

2.6.1 熵理論與共同資訊

早期資訊理論主要用以解決資料傳輸上的資訊計量分析工作，資料從資訊源到目的地的資料傳輸過程中，每一步驟均可能受到不可控制的因素影響，使資料的傳輸過程發生錯誤。為維護資料及分析工作必須建立適當的指標，以衡量資料傳輸的資訊量多寡。一般常用的資訊衡量指標有三種【31】：熵(Entropy)、共同資訊(Mutual information)與識別力(Discrimination)，以下僅敘述MIFS 應用的前二項指標。

若隨機變數 X 對系統產出有影響力，則將 X 視為一個訊息通道(Message Channel)。若變數 X 的值域可等分成 m 個區間，則任一區間 X_k 發生的機率為 p_k 。當 $X=X_k$ 發生，熵值之定義如下：

$$I(x_k) = \log\left(\frac{1}{p_k}\right) = -\log p_k \quad (\text{公式2.3})$$

事件 $X=X_k$ 尚未發生前， $I(x_k)$ 可解釋為系統輸出的不確定(Uncertainty)；當事件發生後，系輸出的不確定性已消失，則 $I(x_k)$ 解釋為該事件發生後的資訊獲得 (Information gain)。 $I(x_k)$ 具有下列物理特性【32】：

1. $I(x_k) = 0, \text{ for } p_k = 1$

若某一事件 X_k 的發生機率 p_k 為1，則無任何不確定性可言，因此無法從該事件的發生獲得任何的資訊。

2. $I(x_k) \geq 0, \text{ for } 0 \leq p_k \leq 1$

若某一事件 X_k 發生，若非因而獲得資訊(>0)，就是沒有獲得任何資訊($=0$)，但絕不會造成資訊損失。

3. $I(x_k) > I(x_j), \text{ for } p_k < p_j$

某一事件發生的機率愈低，一但該事件發生，則觀察者所獲得的資訊愈多。

$I(X_k)$ 的大小視機率值 p_k 而定，隨機變數 X 的熵值之期望值 $H(X)$ (公式

2.4)。若有兩隨機變數 X 與 Y ，則系統的條件熵值之期望值為 $H(X|Y)$ (公式 2.5)。 $H(X|Y)$ 表示觀察到變數 Y 值後，系統對 X 輸出所剩餘的不確定性。 $H(X)-H(X|Y)$ 為透過觀察系統輸出 Y 值後，所降低的不確定性。 $I(X,Y)$ 為二變數 X 與 Y 之間共同資訊計算方式，如(公式2.6)。

$$H(X)=E[I(x_k)]=\sum_k p_k I(x_k)=-\sum_k p_k \log p_k \quad (\text{公式2.4})$$

$$H(X|Y)=H(X,Y)-H(Y) \quad (\text{公式2.5})$$

$$I(X;Y)=H(X)-H(X|Y) \quad (\text{公式2.6})$$

當二隨機變數 X 與 Y 的相關性愈大，二者的共同資訊量 $I(X;Y)$ 愈高，透過其中一者的觀察可降低另一變數輸出值的不確定性。

$I(X;Y)$ 具有下列物理特性【34】：

1. X 與 Y 之共同資訊具有對稱性 (Symmetric)，即 $I(X;Y)=I(Y;X)$
2. X 與 Y 之共同資訊不為負值 (Nonnegative)，即 $I(X;Y) \geq 0$
3. X 與 Y 之共同資訊可以 Y 之熵值表成下式

$$I(X;Y)=H(Y)-H(Y|X) \quad (\text{公式2.7})$$

類神經網路將輸入向量轉換成輸出向量的過程，可視為一個訊息轉換的過程。將每一特徵項目視為訊息輸入通道，而類神經網路的判定結果為系統輸出通道。因此，上述資訊理論 (熵理論與共同資訊) 可應用於評估特徵項目的選擇。

第三章 研究方法與系統架構

3.1 資料範圍

本研究應用的心血管疾病健診資料來源為某國外醫療研究機構(STULONG)1990年至2000年間，關於10年持續的心血管疾病的危險原素縱向研究的數據集合，項目包含了生物檢查以及生活習慣。為了一般國人使用方便，經由心血管疾病專家擷取一般建診常見的生物檢查項目以及生活習慣項目。

3.2 研究工具

本研究的資料前處理步驟應用Microsoft Excel 2003 與Microsoft Access 軟體。將處理過的資料輸入資料庫中，再匯入MATLAB 做為各項研究步驟的輸入資料。應用MATLAB 應用軟體處理屬性挑選及心血管疾病分類之運算。建立類神經網路是利用MATLAB 的類神經網路工具箱 (Neural Network Toolbox)，建構與訓練網路參數的權重值。MATLAB 應用軟體為一套應用於數值計算、數據視覺化及動態模擬的軟體。而且具有簡易的程式語法結構，本研究以此語法進行屬性挑選及心血管疾病分類法則的程式撰寫。

3.3 研究機制設計

本研究之研究機制設計分為建構類神經分類器，建立病患健診數據及生活習慣特徵對於心血管疾病類別之間的關聯性，藉由類神經分類器加以分類。分為兩個階段：(1)篩選病患特徵 (2)建構類神經分類器。

(1)篩選病患特徵

病患特徵篩選採用兩種方式，分別為領域專家之經驗與共同資訊特徵選擇法(MIFS)。結合兩者挑選的屬性作為分類器之輸入變數。

(2)建構類神經網路

以訓練資料集做為訓練網路之資料來源。篩選後的病患健診特徵項目為類神經分類器的輸入神經元(Input Neuron)，而病患所屬心血管疾病類別為輸出神經元(Output Neuron)。完成類神經網路訓練後，類神經分類器就可被應用於預測病患所屬心血管疾病類別。分類器之評估，則進行分類器模擬測試。以測試資料集為網路模擬之輸入資料，所得分類器之準確率做為評估分類器之考量依據。本研究選取最適類神經網路模型做為系統推理機之演算核心。

圖3.1 為本研究BPN網路研究機制設計流程，圖3.2為本研究BPN網路之資料流程圖。

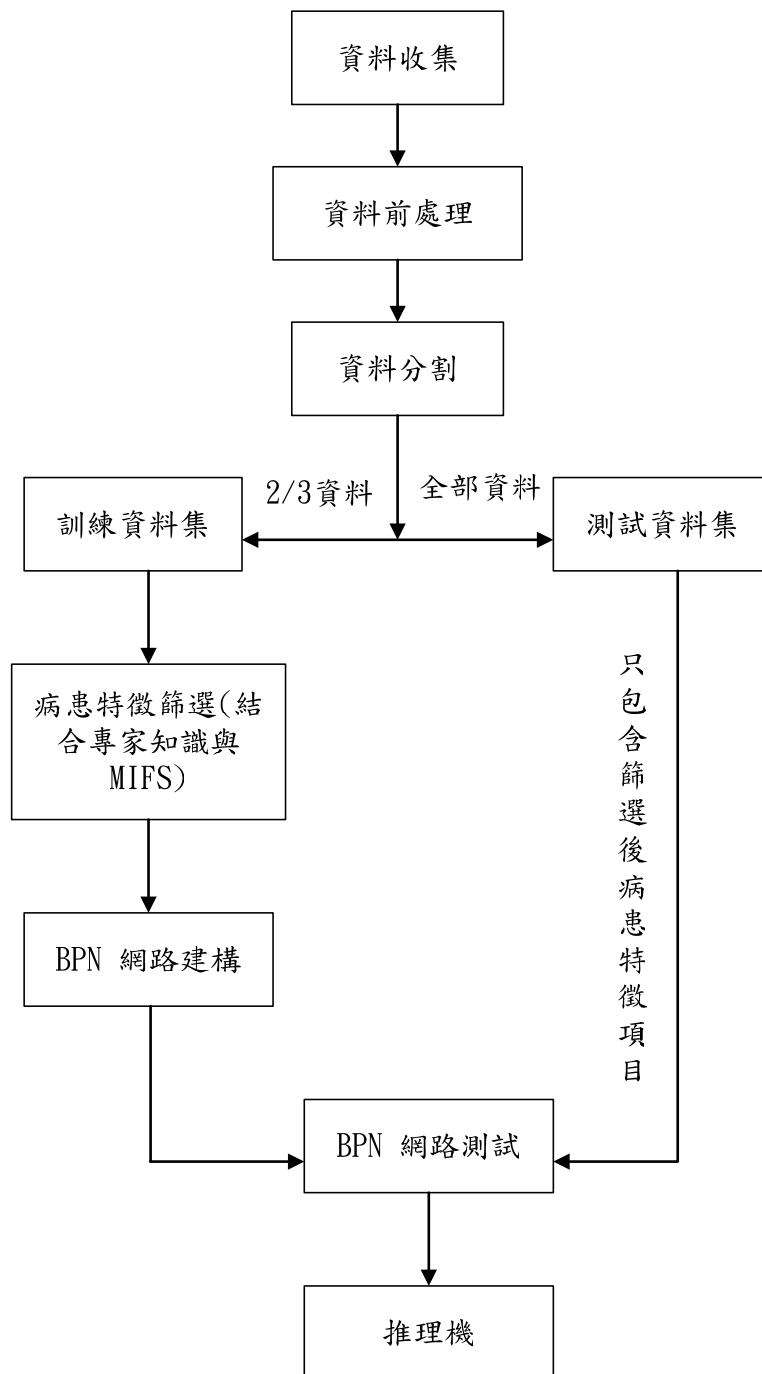


圖 3.1 BPN 網路研究機制設計

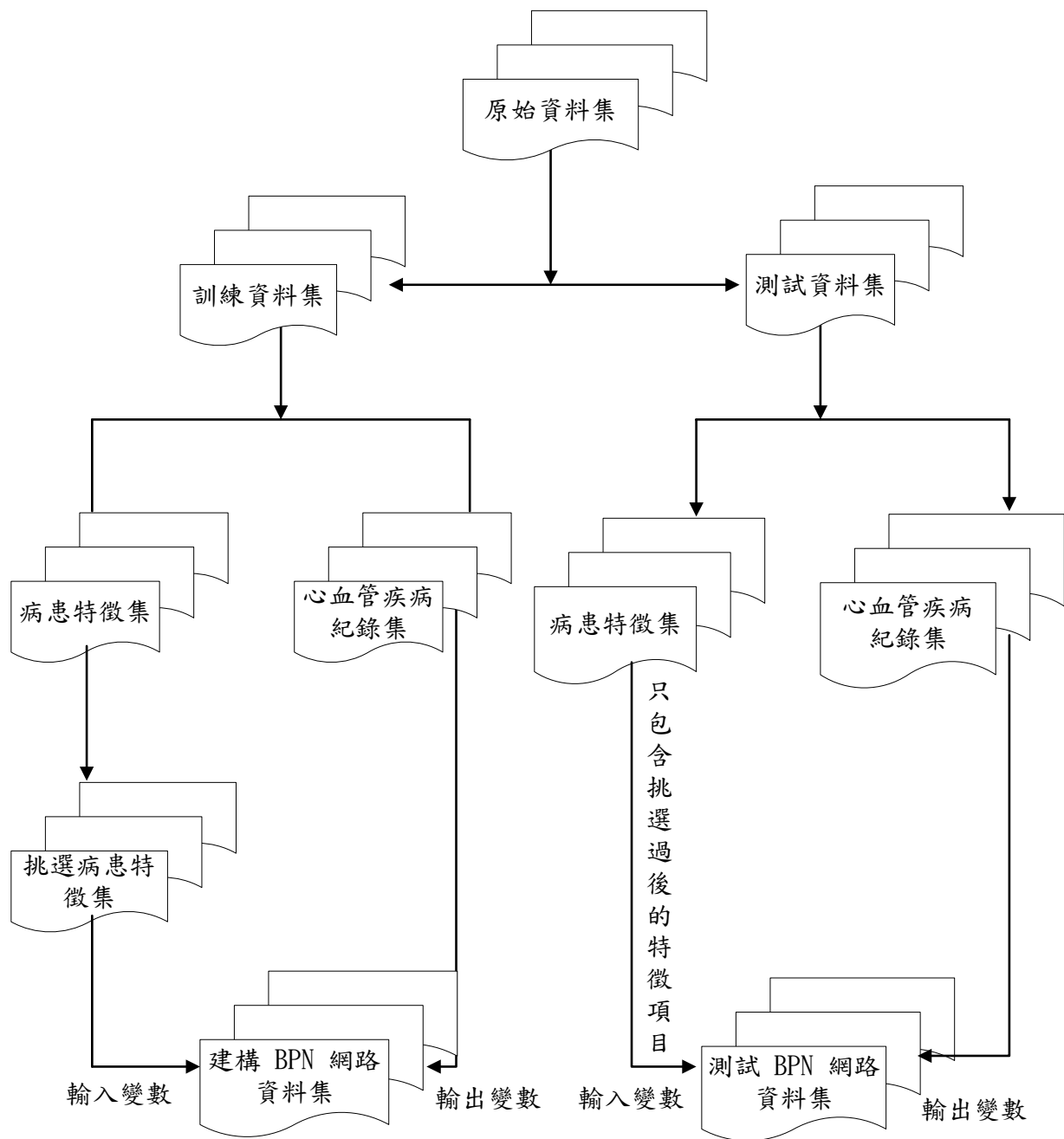


圖 3.2 BPN 網路之資料流程圖

3.3.1 資料搜集

定義資料集範圍及搜集樣本資料。本研究之資料搜集對象為心血管疾病患者健診資料。以某國外心血管疾病研究中心提供於1990年至2000年之心血管疾病患者建診資料為研究對象。研究資料數量：有916份病例記錄。因本研究的資料需求，將原始資料集分為兩個子資料集：病患健診特徵集、心血管疾病記錄集。病患健診特徵集應用於屬性挑選，病患健診特徵集與心血管疾病記錄集應用於建立網路模型與測試。

3.3.2 資料集屬性挑選

病患特徵篩選的資料為原始病患特徵資料集與病患所屬心血管疾病類別。本研究採用兩種方式挑選較小的病患特徵項目集，其足以描述問題。藉由縮小特徵項目資料集的範圍，降低輸入神經元數量，以提高建立類神經分類器的效率。

方法一，由領域知識專家進行特徵項目篩選。

方法二，應用共同資訊特徵選擇法【19】。

MIFS【30】以資訊理論中的熵(Entropy)與共同資訊(Mutual Information)理論為計算基準，評估原始資料集各特徵項目對於判定新血管疾病類別的相對重要程度。MIFS 演算法有兩個參數值，分別為： S ，表所欲挑選的特徵項目個數； β ，表重視未選的特徵項目與已選的特徵項目之間的重複性資訊程度。與一般共同資訊理論不同之處，在於考量特徵項目之間的資訊重複性程度。 β 值愈大，顯示對特徵值間的資訊重複性愈重視。若某特徵項目的不確定性強，且與已選定的特徵項目間資訊重複性低，則此項目重要性序位較高。反之，此特徵項目的重要性序位就較低。

由於本研究特徵項目的屬性值並非布林值(Boolean)，須將病患特徵資料集之各項特徵值重新定義為MIFS 的一個特徵項目。舉例說明：特徵項目「年齡」有五個屬性值：(1,2,3,4,5)，將其轉換為特徵項目1(表示年齡類別1)；特徵項目2(表示年齡類別2)；特徵項目3(表示年齡類別3)；特徵項目4(表示年齡類

別4)；特徵項目5(表示年齡類別5)。結合以上兩種方法所得之特徵項目集，達到資料面與領域知識之整合。整合兩者之可能方式分為下列二種：

(1) MIFS 挑選之特徵項目集合屬於專家挑選之特徵項目集合，則選取兩者之交集集合做為分類器之輸入變數。因MIFS 主要之功能為減化描述問題所需的特徵項目數量，以降低網路學習的複雜度。因此挑選交集集合為分類器之輸入變數。

(2) 挑選之特徵項目集未被完整包含於專家挑選之特徵項目集，即兩者出現差集。此時則採用兩者之聯集集合為分類器之輸入變數。表3.1為原始資料健診與生活習慣項目整理。

表 3.1 原始資料項目如下

編號	欄位名稱	編號	欄位名稱
1	編號	2	年齡
3	地區	4	性別
5	身高	6	體重
7	舒張壓	8	收縮壓
9	總膽固醇	10	高密度膽固醇
11	低密度膽固醇	12	尿酸
13	尿素氮	14	血醣
15	三酸甘油脂	16	煙
17	酒	18	茶
19	咖啡	20	糖
21	疾病		

3.3.3 原始資料前處理

本研究為簡化演算法的計算程序及提高演算效率，設計所需的資料模式。為滿足研究流程中研究方法所需資料型式，本研究之資料前處理部分如下所述：

常用的非線性轉換函數有雙彎曲函數（Sigmoid Function），與雙曲線正切函數（Hyperbolic Tangent Function）。雙彎曲函數之值域在 $(0,1)$ 之間，適用於數據皆為正數時使用；雙曲線正切函數之值域在 $(-1,1)$ 之間，適用於數據若有正號與負號時使用。兩者的共通性為當自變數趨於正負無限大時，函數值趨於常數，其函數圖形分別如圖3.3、圖3.4 表示。

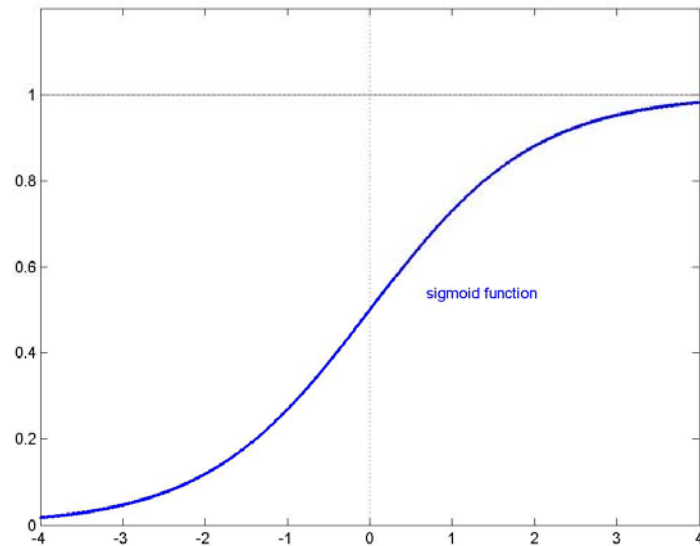


圖 3.3 雙彎曲函數圖

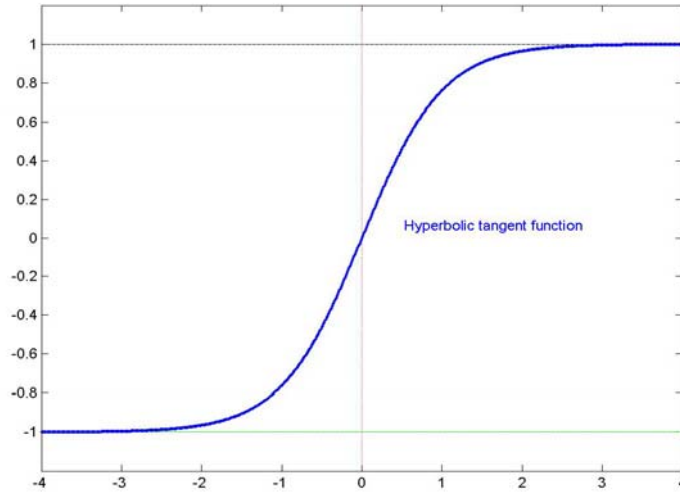


圖 3.4 雙彎曲正切函數圖

由於在BPN網路中，本研究所使用的轉換函數(Transfer Function)為非分線性的雙彎取函數(Sigmoid Function, $f(x) = \frac{1}{1+e^{-x}}$), 其輸出值域為 $[0,1]$, 之間。因此必須將所得到的原始數值資料轉換為 $[0,1]$ 區間的數值, 才能輸入BPN網路中訓練與測試。

於是把原始資料的平均值加上2.5倍原始資料的標準差, 所得到的值轉換為0.9。而把原始資料的平均值減去2.5倍序列的標準差, 所得到的值轉換為0.1。整各數值的轉換方法可用下式說明

$$Nt = \frac{[Xt - (\mu - 2.5 \times \sigma)]}{(\mu + 2.5 \times \sigma) - (\mu - 2.5 \times \sigma)} \times (0.9 - 0.1)$$

Nt 為轉換後的新數值

Xt 為原始資料數值

μ 為原始資料的平均數, $E(Xt) = \mu = 0$

σ 為原始資料的標準差

3.3.4 建構類神經分類器及模式評估

應用MATLAB 應用軟體之類神經網路工具箱，訓練、模擬類神經網路。將選定的病患健診特徵做為訓練類神經網路的輸入變數，而網路的輸出變為病患所患心血管疾病類別。圖3.5為本研究BPN網路分類示意圖。



圖 3.5 BPM 網路分類示意圖

首先，以訓練資料集訓練倒傳遞類神經網路之偏權值。建立分類器需選擇網路模型、神經元之個數 (輸入元、輸出元與隱藏元)。再決定輸入資料集及目標資料集、學習演算法的類型、訓練批次、學習參數。

分類器模擬測試。以測試輸出值與目標輸出值的均方誤差判斷分類器之準確率與評估模擬結果之合理性。測試分類器之輸入資料為測試資料集的病患健診特徵(選取篩選後的特徵項目，為輸入變數)與做為目標輸出值的病患心血管疾病類別。當分類器之測試輸出值與目標輸出值的均方誤差值大於0.5 時，判定此筆記錄為分類錯誤點。分類器之準確率為樣本中分類正確之案件數除以所有案件數【27】。

3.4 倒傳遞網路運作過程

倒傳遞網路的運作過程，分為學習與回想兩部分。其中學習過程是一種監督式學習，它從問題領域中取得訓練範例及目標輸出值，並將訓練範例輸入網路中，利用最陡坡降法 (The Gradient Steepest Descent Method) 反覆地調

節網路的连接加權值及偏權值。倒傳遞網路演算法，分為學習與回想兩部分，茲將演算法的學習步驟及回想步驟詳細如下所示【2】：

1.學習演算法

倒傳遞網路學習演算法共分為8個步驟，說明如下：

步驟1：決定網路的層數及各層間神經元數目。

說明：在此假設網路的架構為輸入層、一層隱藏層級輸出層，級假設網路是三層網路架構，且假設輸入層的神經元數目有 N_{inp} 個、隱藏層的神經元數目有 N_{hid} 個，而輸出層的神經元數目有 N_{out} 個

步驟2：以均佈隨機亂數設定網路的初始加權值及初始偏權值。

說明：由於有 N_{inp} 個輸入神經元與 N_{hid} 個隱藏神經元，所以我們可以用一個雙層迴圈來設定所有輸入層與隱藏層間的初始加權值，方法如下：

$$\begin{aligned} & \text{for } i=1 \text{ to } N_{inp} \\ & \text{for } h=1 \text{ to } N_{hid} \\ & W_{xh}[i][h]=\text{均佈隨機亂數} \end{aligned}$$

如果我們令 $W_{xh}[i][h]$ 為隱藏層第 h 個神經元與輸出層第 j 個神經元間的加權值，則我們設定所有隱藏層與輸出層間的初始加權值的方法如下：

$$\begin{aligned} & \text{for } h=1 \text{ to } N_{hid} \\ & \text{for } j=1 \text{ to } N_{out} \\ & W_{hy}[h][j]=\text{均佈隨機亂數} \end{aligned}$$

再來我們要設定網路中的初始偏權值，要注意的是只有隱藏層級輸出層才有偏權值，輸入層是沒有的。事實上，輸入層是沒有运算能力的，它只是將一個神經元接收到的訊號平行輸出至隱藏層各個神經元中。若令 $\theta_h[h]$ 為隱藏層第 h 的神經元的偏權值， $\theta_y[j]$ 為輸出層第 j 個神經元的偏權值，則設定初始偏權值的方法如下：

$for\ h=1\ to\ N_{hid}$
 $\theta\ h[h]=$ 均佈隨機亂數
 $for\ j=1\ to\ N_{out}$
 $\theta\ y[j]=$ 均佈隨機亂數

步驟 3：輸入訓練樣本 $x[1], x[2], \dots, x[N_{inp}]$ 及目標輸出值 $T[1], T[2], \dots, T[N_{out}]$

說明：輸入值 $x[1], x[2], \dots, x[N_{inp}]$ 可為任意的實數值，但是由於倒傳遞網路採用雙彎曲函數（sigmoid function）當作神經元的非線性轉換函數，網路的推論輸出值的值域為 $(0, 1)$ 之間，所以目標輸出值 $T[1], T[2], \dots, T[N_{out}]$ 其值域也必須落在 $(0, 1)$ 之間。

步驟 4：計算網路的推論輸出值 $y[1], y[2], \dots, y[N_{out}]$ 說明：計算的方法是先算出隱藏層的輸出值，方法如下：

$for\ h=1\ to\ N_{hid}$
 $net\ h[h]=\sum_{i=1}^{N_{inp}} W_{xh}[i][h]x[i]-\theta\ h[h]$
 $for\ h=1\ to\ N_{hid}$
 $H[h]=\frac{1}{1+e^{-net\ h[h]}}$

其中 $net\ h[h]$ 為隱藏層第 h 個神經元的加權乘積和，而 $h[h]$ 為隱藏層第 h 個神經元的輸出值，它將收集到的加權乘積和 $net\ h[h]$ 再作一次非線性轉換。由於輸出值的輸入訊號來自隱藏層的輸出值，所以其推論輸出值可計算如下：

$for\ j=1\ to\ N_{out}$
 $net\ y[j]=\sum_{h=1}^{N_{hid}} W_{hy}[h][j]H[h]-\theta\ y[j]$
 $for\ j=1\ to\ N_{out}$
 $Y[j]=\frac{1}{1+e^{-net\ y[j]}}$

其中 $net\ y[j]$ 及 $Y[j]$ 分別是輸出層第 j 個神經元的加權乘積和及推論輸出

值。

步驟5：計算輸出層與隱藏層的差距量。

說明：計算輸出層差距量的公式如下：

$$\begin{aligned} & \text{for } j=1 \text{ to } N_{out} \\ & \delta y[j] = (1 - Y[j])(T[j] - Y[j]) \end{aligned} \quad (\text{公式3.1})$$

其中 $\delta y[j]$ 是輸出層第 j 個神經元的差距量。在公式3.1 中， $T[j] - Y[j]$ 表示目標輸出值與網路推論輸出值間的誤差，所以 $\delta y[j]$ 表示 $T[j]$ 與 $Y[j]$ 之間的誤差量度。而計算隱藏層差距量的公式如下：

$$\begin{aligned} & \text{for } h = 1 \text{ to } N_{hid} \\ & \delta h[h] = H[h](1 - H[h]) \sum_{j=1}^{N_{out}} W_{hy}[h][j] \delta y[j] \end{aligned} \quad (\text{公式3.2})$$

其中 $\delta h[h]$ 表示隱藏層第 h 個神經元的差距量。請注意，在公式3.2 中包含了一項子式： $\sum_{j=1}^{N_{out}} W_{hy}[h][j] \delta y[j]$

此式表示輸出層差距量的加權乘積和。所以 $\delta h[h]$ 的計算與輸出層的差距量有關，這意味著我們將輸出層的誤差倒傳至隱藏層來計算其差距量，這是此網路之所以有「倒傳遞」名稱的由來。

步驟6：計算各層間的加權值修正量及偏權值修正量。

說明：若令 $\Delta W_{hy}[h][j]$ 表示隱藏層第 h 個神經元與輸出層第 j 個神經元的加權值修正量，且令 $\Delta \theta y[j]$ 表示輸出層第 j 個神經元的偏權值修正量，則計算其間所有加權值及偏權值修正量的方法如下：

$$\begin{aligned} & \text{for } j=1 \text{ to } N_{out} \\ & \Delta W_{hy}[h][j] = \eta \delta y[j] H[h] \end{aligned} \quad (\text{公式3.3})$$

$$\begin{aligned} & \text{for } j=1 \text{ to } N_{out} \\ & \Delta \theta y[j] = -\eta \delta y[j] \end{aligned} \quad (\text{公式3.4})$$

其中 η 為學習速率，一般取值為 0.1~1.0。有時為了加速網路的收斂速度，可將公式3.3 與 3.4 改寫成

$$\Delta W_{hy}[h][j] = \eta \delta y[j] H[h] + \alpha \Delta W_{hy}[h][j] \quad (\text{公式3.5})$$

$$\Delta \theta y[j] = -\eta \delta y[j] + \alpha \theta y[j] \quad (\text{公式3.6})$$

其中 α 為慣性因子，一般取值為 0.0~0.9。同理，若令 $\Delta W_{xy}[i][h]$ 為輸入層第 i 個神經元與隱藏層第 h 個神經元間有加權修正量，令 $\theta h[h]$ 為隱藏層第 h 個神經元間的偏權修正量，則計算其間所有加權值及偏權值修正量的公式如下：

$$\begin{aligned} & \text{for } i = 1 \text{ to } Ninp \\ & \quad \text{for } h = 1 \text{ to } Nhid \\ & \quad \quad \Delta W_{xh}[i][h] = \eta \delta h[h] x[i] + \alpha \Delta W_{xh}[i][h] \end{aligned} \quad (\text{公式3.7})$$

$$\begin{aligned} & \text{for } h = 1 \text{ to } Nhid \\ & \quad \Delta \theta h[h] = -\eta \delta h[h] + \alpha \Delta \theta h[h] \end{aligned} \quad (\text{公式3.8})$$

步驟7：更新各層間有加權值及偏權值。

說明：更新隱藏層與輸出層間有加權值及輸出層偏權值的方法如下：

$$\begin{aligned} & \text{for } h = 1 \text{ to } Nhid \\ & \quad \text{for } j = 1 \text{ to } Nout \\ & \quad \quad W_{hy}[h][j] = W_{hy}[h][j] + \Delta W_{hy}[h][j] \end{aligned} \quad (\text{公式3.9})$$

$$\begin{aligned} & \text{for } j = 1 \text{ to } Nout \\ & \quad \theta y[j] = \theta y[j] + \Delta \theta y[j] \end{aligned} \quad (\text{公式3.10})$$

同理，更新輸入層與隱藏層間有加權值及隱藏層偏權值的方法如下：

$$\begin{aligned} & \text{for } i = 1 \text{ to } Ninp \\ & \quad \text{for } h = 1 \text{ to } Nhid \\ & \quad \quad W_{xh}[i][h] = W_{xh}[i][h] + \Delta W_{xh}[i][h] \end{aligned} \quad (\text{公式3.11})$$

$$\begin{aligned} & \text{for } h = 1 \text{ to } Nhid \\ & \quad \theta h[h] = \theta h[h] + \Delta \theta h[h] \end{aligned} \quad (\text{公式3.12})$$

步驟8：重複步驟3 至步驟7，直至網路收斂。

說明：學習過程通常以一次一個訓練樣本進行，直到網路學習完所有的

訓練樣本，稱為一個學習循環（Learning Circle），我們可以讓網路重複學習數個學習循環，直至網路收斂為止。為了測試網路是否收斂，我們定義下列誤差函數來表示網路的學習品質：

$$E = \left(\frac{1}{2}\right) \sum_j (T[j] - Y[j])^2 \quad (\text{公式3.13})$$

此式表示輸出層各個神經元的平方誤差和。因為在學習過程中，我們希望網路的推論輸出值 $Y[j]$ 與目標輸出值 $T[j]$ 越接近越好，所以公式3.13的計算值應小於一個合理的範圍才行。

2. 回想演算法

倒傳遞網路回想演算法，其步驟說明如下：

步驟1：設定網路的層數及各層間神經元的數目。

步驟2：讀入已訓練好的網路加權值及偏權值。

步驟3：輸入一個測試範例 $x[1], x[2], \dots, x[Ninp]$ 。

步驟4：計算網路的推論輸出值 $y[1], y[2], \dots, y[Nout]$

說明：此步驟與學習演算法的步驟4 相同。

3.4.1 BPN 網路參數設定

A. 輸入變數(Input Variable)

將經由資料集屬性挑選後的欄位數值作為輸入變數的資料來源。輸入變數為17個單元。

B. 隱藏層數(Hidden Layer)

隱藏層數越多，網路學習時間也會耗費較多。Villiers研究指出【48】，隱藏層為一層或兩層時之表現皆類似，但因兩層隱藏層較易陷入局部最小值。根據Castelli之研究顯示【33】，包含單一隱藏層之網路結構，足以描述複雜之非線性關係，且可達到良好之正確度，因此本研究中隱藏層數採

用一層。圖3.6 為本研究單階隱藏元網路架構。

C. 隱藏層處理單元數(Process Element of Hidden Layer)

Katz指出隱藏層單元數目=1/2(輸入層單元數)至3×(輸入層單元數)之間【41】。本研究隱藏層處理單元數目則選定3×(輸入層單元數目)。

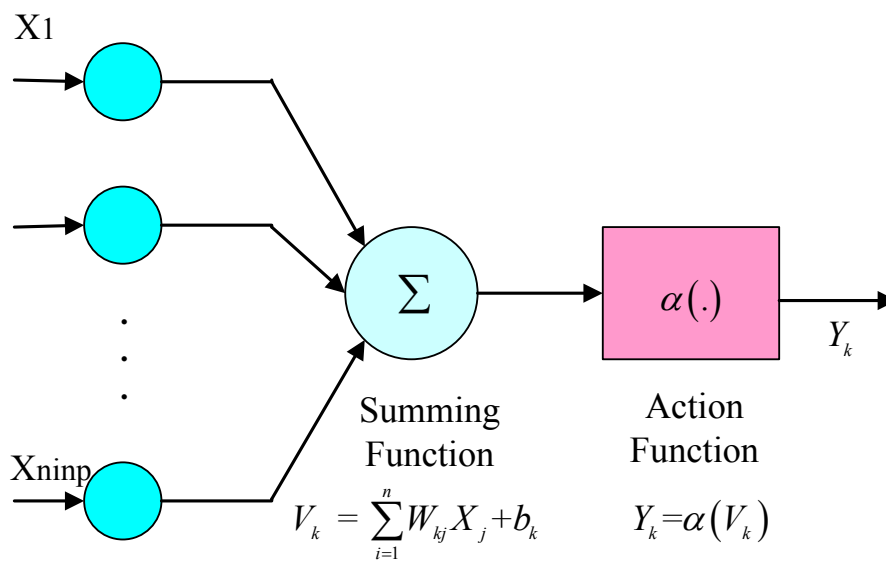


圖 3.6 單階隱藏元網路架構

D. 轉換函數(Transfer Function)

本研究架構所採用的轉換函數為非分線性的雙彎取函數(Sigmoid Function)，函式如下

$$f(x) = \frac{1}{1+e^{-x}} \quad (\text{公式3.13})$$

E. 學習速率(Learning Rate)

較大的網路加權值修正量，可以較快逼近函數最小值，但過大的學習速率將導致網路加權值修正過量，造成數值震盪而難以達到收斂的目的。本研究架構中的BPN網路中，控制每次以最陡坡降法(The Gradient Steepest Descent Method)，最小化誤差函數的步幅，為網路的學習速率。

本研究將初始學習速率設為0.7，學習速率的遞減函數為0.95。每經100、

300、700、1500、3100、...、 $100 \times (1+2^1+\dots+2^{n-1})$ 個學習循環，學習速率會遞減一次。當學習循環數 ≤ 100 循環時，學習速率為 0.7。當 $100 \leq$ 學習循環數 ≤ 300 ，學習速率為 $0.7 \times (0.95) = 0.665$ ，依此類推。所以本研究架構的網路學習速率如表 3.2 所示：

表 3.2 BPN 網路學習速率遞減之情形

學習循環數	100	300	700	...	$100 * (1+2^1+\dots+2^{n-1})$
學習速率	0.7	$0.7 \times (0.95)$	$0.7 \times (0.95)^2$...	$0.7 \times (0.95)^{n-1}$

F. 輸出變數(Output Variable)

由於系統主要功能為預測所罹患的心血管疾病類別，因此輸出層的輸出單元設定為 14 個。

G. 慣性因子(Momentum Factor)

即加上某比例的上次加權值改變量，慣性因子的作用為改善網路加權值在收斂過程中震盪的現象及加速收斂，慣性項可以是改善學習速度與學習精度的方法之一。在本研究中設定為 0.5。

H. 初始加權值(Initial Weight)

由於研究開始進行前對於健診項目與心血管疾病的相關性無法得知。本研究將初始加權值設為 1 至 -1 之間的隨機均部亂數。

I. 學習法則(Learning Rule)

本研究採用 Delta-Rule 為 BPN 網路學習法則。當網路每學習一個範例，網路連結加權值便更新一次。

3.4.2 BPN 神經網路學習與測試

學習過程適當的學習次數可以讓網路有比較好的歸納能力，以求在測試過程當網路遇到沒有學習過的樣本時，能準確地預測。而過度的訓練次數會使網路產生過度學習的問題，使網路產生記憶現象。因此，決定網路學習程度的方法有兩種：1.決定學習的次數 2.決定誤差的範圍。但兩者都有其缺點，如：決定學習的次數過小時則網路不會收斂；若一味追求最小誤差，可能會導致過度訓練 (over-training) 的問題，而使得網路的學習誤差值極小但測試誤差可能變大【27】。

為了避免學習不足(lack of fit)與學習過度(over-fitting)的情形，希望找出適當的學習循環次數，並經由總錯率的檢驗找出最適合系統使用的BPN網路模型。因此本研究設定了三種不同的BPN學習循環次數分析度(Resolution)，其目的是為了使不同學習循環次數下，對於BPN網路的學習績效能夠更有效的了解。以Resolution 1、Resolution 2、Resolution 3表示，其說明如下：

Resolution 1:

設定BPN網路的學習循環次數為1000次。當網路每學習10次循環，便記錄一次網路學習範例與測試範例之網路輸出值以便分析其學習績效，共100組網路輸出值的分析結果。

Resolution 2:

設定BPN網路的學習循環次數為10000次。當網路每學習100次循環，便記錄一次網路學習範例與測試範例之網路輸出值以便分析其學習績效，共100組網路輸出值的分析結果。

Resolution 3:

設定BPN網路的學習循環次數為50000次。當網路每學習500次循環，便記錄一次網路學習範例與測試範例之網路輸出值以便分析其學習績效，共100組網路輸出值的分析結果。

以上三組Resolution試驗中，分別會有一百組BPN網路在不同學習循環次數下，所得到的網路輸出值分析結果，藉由分析三組試驗找出每各組別中最適合的學習循環次數。

3.4.3 預測準確度的評估標準

1. 誤差均方根(RMS Error)

$$\text{誤差均方根} = \sqrt{\frac{\sum_{P=1}^M \sum_{j=1}^N (T_j^P - Y_j^P)^2}{M \times N}}$$

T_j^P = 第P個範例的第j各輸出單元之目標輸出值

Y_j^P = 第P個範例的第j各輸出單元之推論輸出值

M = 範例數目

N = 輸出層單元數目

2. 總錯率

總錯率 = 誤分類範例總數 / 範例總數

以範例總數 100 來說，若誤分類範例總數為 20，則總錯率為 20/100=0.2=20%

一般而言，類神經網路在訓練學習過程中，誤差均方根與總錯率收斂情形並不一定相同。總錯率的評估標準，通常是使用在範例只有一個正確的分類的時候；而預測結果為數值者，則誤差函數及誤差均方根較為適用。

在BPN網路模式建構時因心血管疾病都已經加以編碼，所以使用誤差均方根來判斷最佳學習循環的次數。當BPN網路模式建構完成時，輸出值為心血管疾病的分類，所以使用總錯率來判斷最佳BPN網路模式，選取最佳BPN網路模式作為系統分類的依據。

3.5 系統架構

本研究可將系統架構分成知識庫和推理機兩大元件組成（圖3.7），其中推理機的功能是判斷可分成比對病歷資料庫與利用倒傳遞網路（BPN）來分析預測，且倒傳遞網路是具備有自我學習的能力，可以仿效專家的思考模式可預測所罹患的心血管疾病。系統將輸入條件與病例資料庫的病例比對後來產生結果。若條件與法則無法完全符合時，則由倒傳遞網路（BPN）來處理，其系統流程如圖3.8所示。

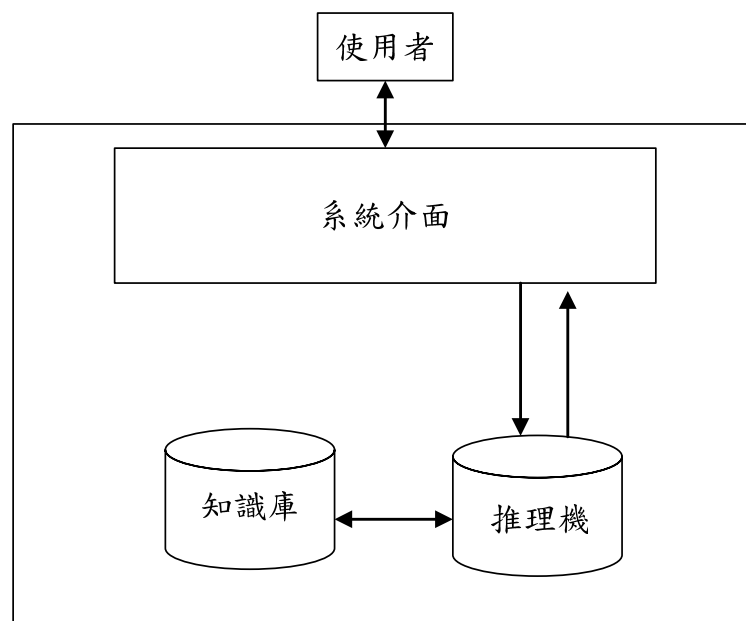


圖 3.7 系統架構圖

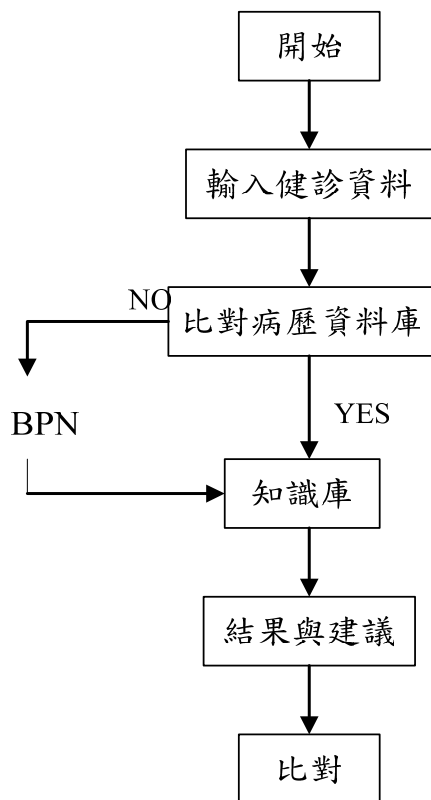


圖 3.8 使用者流程圖

3.5.1 系統主要元件

1. 知識庫 (Knowledge Base)

知識是專家系統的原動力，透過專家和知識工程師的溝通合作，共同將專家知識轉換成知識庫內容【18】，將知識來源分成專業知識和一般知識。專業知識是由心臟科專業醫師所提供對於心血管疾病患者日常生活所提出的建議，而一般知識為系統依機率所產生的資訊。知識庫是系統從病歷資料庫中的資料整理出不重複的法則，每個法則包含狀態(Condition)和行為(Action)的兩個部分，如表 3.8 所示，則健診項目為狀態、心血管疾病類別行為。

2. 推理機 (Inference Engine)

本研究系統的推理機有兩大功能。一是轉換功能，是將病歷資料庫的資料，經過整個轉換處理再存入訓練資料表中。另一功能是比较處理，本系統的比较處理是先將使用者輸入條件與病例資料庫的法則比较後來產生結果。若條件與法則無法完全符合時，則由倒傳遞網路 (BPN) 來處理，利用訓練

資料表運算，運算後取最大加權值。所對應的科別，即為預測分析的心血管疾病類別。

3.類神經網路單元

本系統對於無完全符合病例法則的條件，皆由倒傳遞類神經網路來處理，其處理過程來可分成二個階段：學習過程（Learning）和回想過程（Recalling），其運作流程如圖 3.5 和圖 3.6 所示。本研究的倒傳遞神經網路，具有 17 個輸入單元，14 個輸出單元為，一層隱藏層，51 個隱藏節點，其學習過程與回想過程的步驟如下所示：

（1）學習步驟

- a.從病歷資料庫中篩選出經由倒傳遞網路預測的資料，因資料無法完全符合病歷資料庫中的法則，將此交由專家來進行修改。
- b.將修改後的資料匯入訓練資料集，新的訓練資料集經由學習演算法來訓練，直至網路收斂，以達到系統自我學習的目的。
- c.輸出已訓練好的網路加權值及偏權值。

（2）回想步驟

- a.讀入已訓練好的網路加權值與偏權值。
- b.輸入沒有完全符合病例法則的使用者輸入條件。
- c.以回想演算法運算取最大加權值所對應的類別，即為預測分析心血管疾病類別。

第四章 BPN 網路實作與系統試驗

4.1 病患特徵篩選

本研究為提高建立分類器之效益及效率，選出對於療程類別具有重要性之病患特徵做為分類器之輸入神經元。本研究同時採用領域知識專家與之共同資訊特徵選擇法【30】挑選病患特徵項目(表4.1)。由領域知識專家挑選的病患特徵項目(表4.2)。扣除病患代碼等欄位，所得病患特徵項目共17項(表4.3)。

應用共同資訊特徵選擇法進行上述病患特徵值篩選【30】提出共同資訊特徵選擇法。以資訊理論中的熵與屬性的共同資訊為計算基準，以評估資料集中各項特徵值對於療程類別的相對重要程度。本研究將上述病患特徵資料集各特徵項目之屬性值定義為MIFS 的一個特徵，如：「年齡」原有5個屬性值(1, 2, 3, 4, 5)，轉換為MIFS 的5個特徵，「年齡類別1」、「年齡類別2」與「年齡類別3」、「年齡類別4」與「年齡類別5」。

MIFS挑選之特徵項目集未被完整包含於專家挑選之特徵項目集，即兩者出現差集。此時則採用兩者之聯集集合為分類器之輸入變數。將MIFS 未選定的特徵值再次轉換成病患特徵資料集之特徵項目，並由資料集中刪去此特徵項目欄位。因此，依據實驗的療程類別所篩選出病患特徵項目集合，需刪去的特徵欄位包含：「就醫序號」、「地區」、「尿素氮」。取MIFS挑選之特徵和心血管疾病專家挑選之特徵之聯集，表4.1 MIFS採用的變數項目。表4.2 專家採用的變數項目。總共挑選出17項病患特徵，挑選出的特徵項目列表，如同表4.3 所示。

表 4.1 MIFS 採用的變數項目

編號	欄位名稱	編號	欄位名稱
1	年齡	2	性別
3	身高	4	體重
5	舒張壓	6	收縮壓
7	總膽固醇	8	高密度膽固醇
9	低密度膽固醇	10	尿酸
11	煙	12	茶
13	咖啡	14	糖

表 4.2 專家採用的變數項目

編號	欄位名稱	編號	欄位名稱
1	年齡	2	性別
3	身高	4	體重
5	舒張壓	6	收縮壓
7	總膽固醇	8	高密度膽固醇
9	低密度膽固醇	10	三酸甘油脂
11	血糖	12	煙
13	酒	14	茶
15	咖啡	16	糖

表 4.3 最終採用的變數項目

編號	欄位名稱	編號	欄位名稱
1	年齡	2	性別
3	身高	4	體重
5	舒張壓	6	收縮壓
7	總膽固醇	8	高密度膽固醇
9	低密度膽固醇	10	尿酸
11	三酸甘油脂	12	血糖
13	煙	14	酒
15	茶	16	咖啡
17	糖		

4.2 類神經分類器訓練與模擬

本研究採用類神經分類器進行病患療程分類預測。將病患特徵項目做為分類器的輸入變數，心血管疾病為目標輸出變數，以訓練類神經分類器。以4.1節篩選出的17項病患特徵做為類神經分類器的17個輸入變數，隱藏層為一層隱藏單元為51個，輸出層為心臟病類別單元數為14個，學習法則採用Delta-Rule為BPN網路學習法則，最陡坡降法(The Gradient Steepest Descent Method)，最小化誤差函數的步幅，為網路的學習速率，慣性因子設定為0.5。

4.2.1 類神經分類器實做

A. Resolution 1(每10個學習循環紀錄一次BPN網路輸出層輸出值):

我們計算BPN網路推論值(O_i)與輸出目標值(X_i)之 $RMSE_x$ 值，圖4.1為Resolution1中，學習資料集與測試資料集之 $RMSE_x$ 值隨著不同學習循環數下

的變化，X軸的數值表示學習循環次數，Y軸的數值代表 $RMSE_x$ 值。

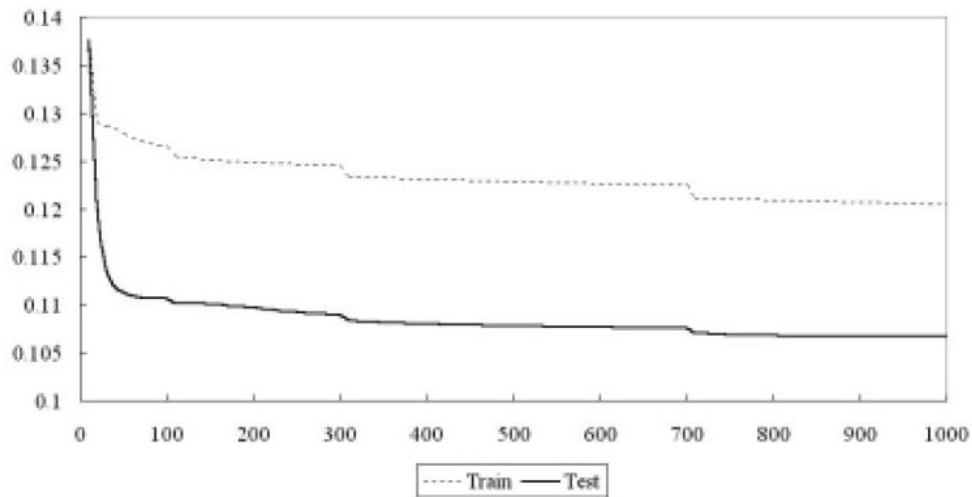


圖 4.1 Resolution1

由圖4.1可以觀察出學習資料集、測試資料集的 $RMSE_x$ 一直持續下降，因此以傳統 $RMSE_x$ 法來判斷，BPN網路在Resolution 1中學習1000個循環數還不夠，仍然必須讓BPN網路繼續學習。

B. Resolution 2(每100個學習循環紀錄一次BPN網路輸出層輸出值):

我們計算BPN網路推論值(O_i)與輸出目標值(X_i)之 $RMSE_x$ 值，圖4.2為Resolution1中，學習資料集與測試資料集之 $RMSE_x$ 值隨著不同學習循環數下的變化，X軸的數值表示學習循環次數，Y軸的數值代表 $RMSE_x$ 值。

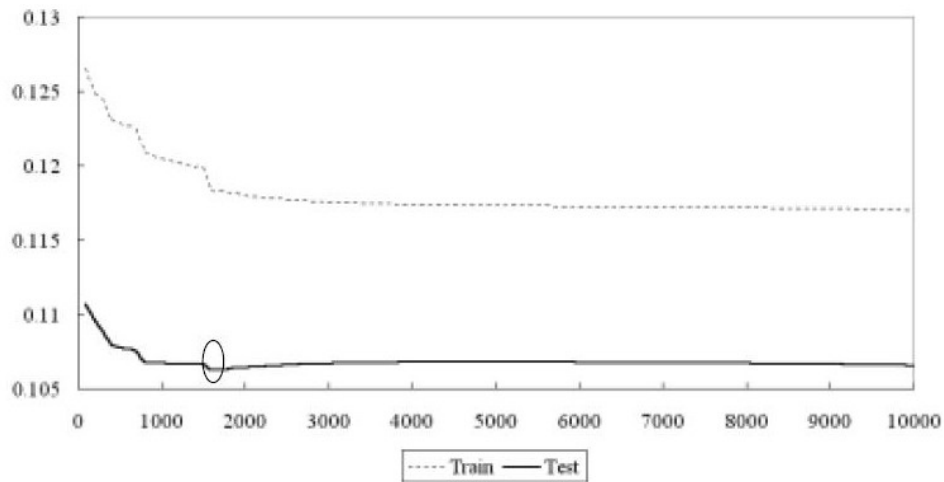


圖 4.2 Resolution2

由圖4.2可以觀察出測試資料集的 $RMSE_x$ 值在約1600個學習循環次數左右其值變化趨勢緩慢且有開始上升之趨勢，因此以傳統 $RMSE_x$ 法來判斷，BPN網路在學習約1600各循環次數左右便可停止。

C. Resolution 3(每500個學習循環紀錄一次BPN網路輸出層輸出值):

我們計算BPN網路推論值(O_t)與輸出目標值(X_t)之 $RMSE_x$ 值，圖4.3為Resolution1中，學習資料集與測試資料集之 $RMSE_x$ 值隨著不同學習循環數下的變化，X軸的數值表示學習循環次數，Y軸的數值代表 $RMSE_x$ 值。

由圖4.3可以觀察出測試資料集的 $RMSE_x$ 值在約17000個學習循環後，其值明顯往上升，而學習資料集的 $RMSE_x$ 值在20500各學習循環後也開始往上升。所以BPN網路在17000各學習循環後已經開始過度學習了。

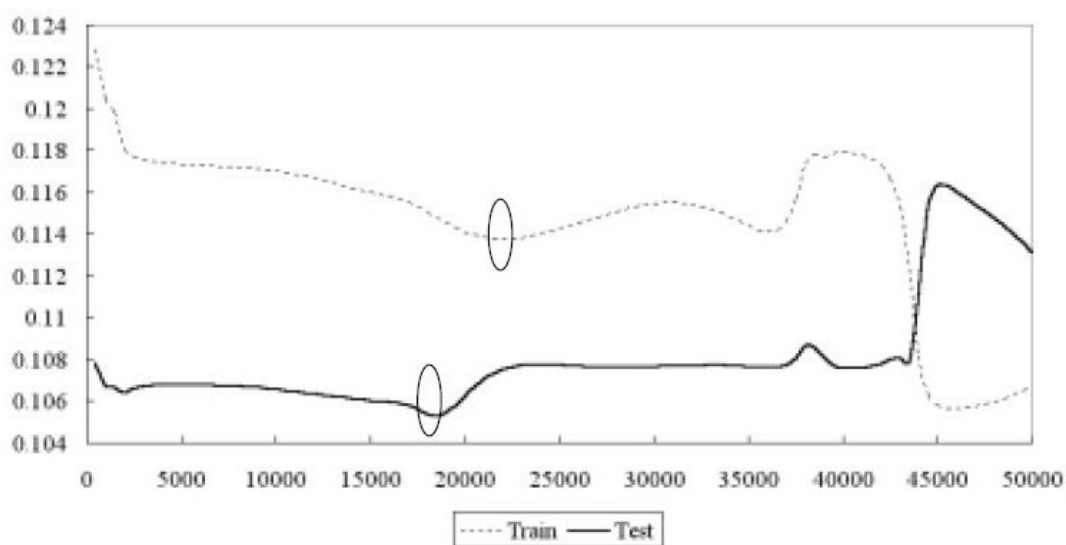


圖 4.3 Resolution3

由以上三個不同的解析度分析，以 $RMSE_x$ 來決定BPN網路適當學習循環次數時，應在1600至17000各學習循環之間，而在Resolution 3中，17000各學習循環次數後，可視為以基開始過度學習。

由於在BPN網路當中，我們輸入網路中訓練的資料集，其輸入變數(Input Variable)及其對應的輸出目標值都是包含雜質的資料。若以輸出層輸出目標值(X_i)與網路推論值(O_i)的 $RMSE_x$ 值來評估BPN網路學習的績效，則當BPN網路值很低時，表示網路推論值(O_i)很接近輸出目標值(X_i)，已經把雜訊給過度學習(over-fitting)到BPN預測模型中。最小的 $RMSE_x$ 值並不表示其擁有最高的鑑別成效，因為 $RMSE_x$ 值所代表意義僅為經由演算法所得之結果，與實際預測後的鑑別率未必成絕對的相關，因此，在選取最佳網路結構及模式時，除了將 $RMSE_x$ 值作為比較依據外，同時也應考量整體鑑別率之優劣。

為求最終網路的實用性與準確性採取上述三種Resolution中個別最佳學習循環次數來做比較。Resolution 1因為 $RMSE_x$ 值並未收斂所以不加以考慮。在Resolution 2中 $RMSE_x$ 出現最低值時學習循環次數為1600時，因此將Resolution 2中學習循環次數為1600的BPN網路列入考慮。在Resolution 3中 $RMSE_x$ 出現最低值時學習循環次數為17000時，因此將Resolution 3中學習循環次數為17000的BPN網路列入考慮。

本研究採用類神經分類器進行些血管疾病分類預測。將病患特徵項目做為分類器的輸入變數，心血管疾病類別為目標輸出變數，以訓練類神經分類器。以4.1節篩選出的17項病患特徵做為類神經分類器的17個輸入變數，本研究應用MATLAB 6.0 版建立與訓練單階隱藏層模型 (Single Hidden Layer model)。進行兩組網路架構之訓練與模擬，相異之處為學習循環次數：分類器一，1600個學習循環，學習率為 $0.7 \times (0.9)^4 = 0.5701$ ，圖4.4為分類器一架構圖；分類器二，17000各學習循環，學習率為 $0.7 \times (0.9)^7 = 0.4888$ ，圖4.5為分類器二架構圖。

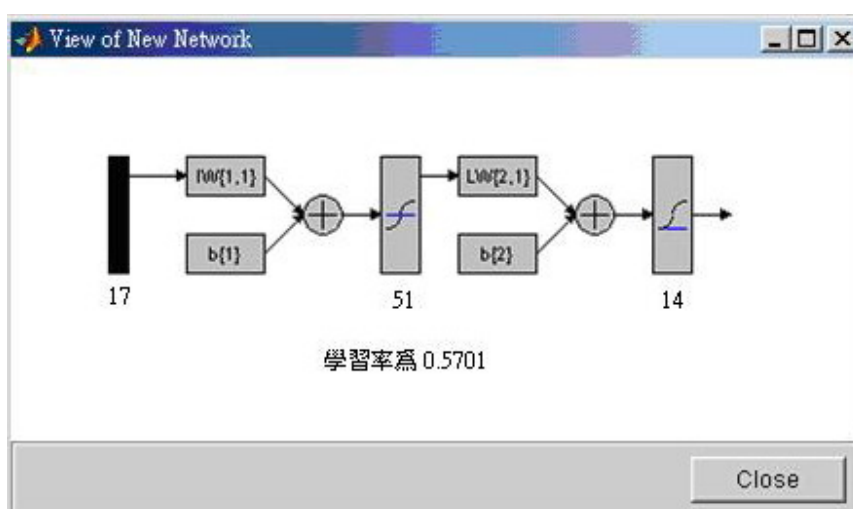


圖 4.4 分類器一

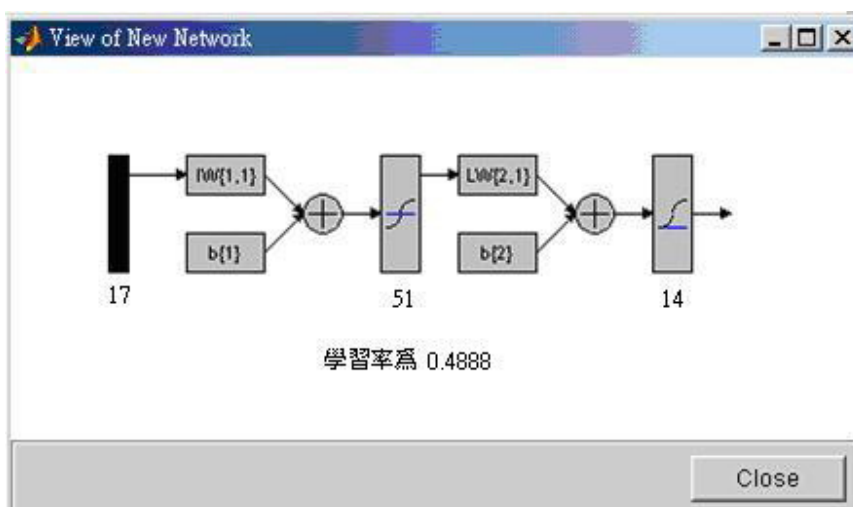


圖 4.5 分類器二

分類器一，17個輸入變數、單一層隱藏層、51 個隱藏元、以最陡坡降法 (The Gradient Steepest Descent Method)，最小化誤差函數的步幅，為網路的學習速率 $0.7 \times (0.95)^{n-1}$ 、14個輸出變數、慣性因子設定為0.5、採用Delta-Rule為BPN網路的學習法則、1600個學習循環。建立網路模型。

初始化網路權數值(Weights)之後，將訓練資料集 (610筆記錄)之病患特徵資料集與所屬心血管疾病類別資料集匯入進行網路練。待網路權數值穩定後，匯入完整資料集 (916筆記錄)的病患特徵與所屬心血管疾病類別資料集完成網路模擬。記錄與分析模擬分類的均方誤差值，藉以評估分類器之準確率。以假設檢定進行分類器之評估。

本研究以誤差值0.5 為標準，若誤差值大於0.5則判定此點分類錯誤。圖4-6，橫座標為病患編號，縱座標為病患之療程分類誤差值。分類器一所得之分類案件數為32 件案例，錯誤率為0.0349。假設顯著水準 $\alpha=0.05$ 時，可接受的分類器之療程分類準確率為96%，則此分類器之療程分類準確率可否滿足。由下列之計算式可知檢定結果為不否定虛無假設 (H_0)，此分類器之分類準確率於97%之標準內。圖4.6 為本研究分類器一誤差統計圖

$$\hat{P} = \frac{32}{916} = 0.0349, \alpha = 0.05$$

$$H_0: P \leq 0.03, H_1: P > 0.03$$

$$Z_0 = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.0349 - 0.03}{\sqrt{\frac{0.97 * 0.03}{916}}} = 0.8703$$

$$Z_0 < Z_{0.95} = 1.645$$

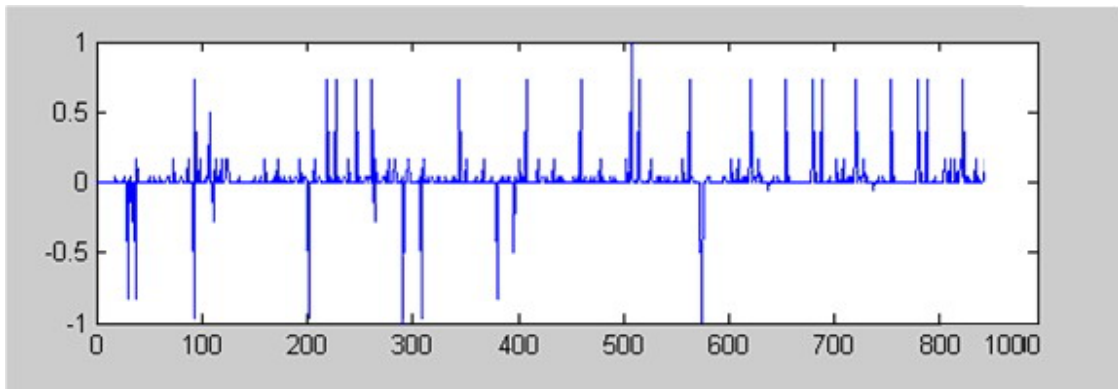


圖 4.6 分類器一誤差統計圖

分類器二，17個輸入變數、單一層隱藏層、51個隱藏元、以最陡坡降法(The Gradient Steepest Descent Method)，最小化誤差函數的步幅，為網路的學習速率 $0.7 \times (0.95)^{n-1}$ 、14 個輸出變數、慣性因子設定為0.5、採用Delta-Rule為BPN網路的學習法則、17000個學習循環。建立網路模型。

初始化網路權數值(Weights)之後，將訓練資料集 (610筆記錄)之病患特徵資料集與所屬心血管疾病類別資料集匯入進行網路練。待網路權數值穩定後，匯入完整資料集 (916筆記錄)的病患特徵與所屬心血管疾病類別資料集完成網路模擬。記錄與分析模擬分類的均方誤差值，藉以評估分類器之準確率。以假設檢定進行分類器之評估。

本研究以誤差值0.5 為標準，若誤差值大於0.5則判定此點分類錯誤。圖4.7，橫座標為病患編號，縱座標為病患之療程分類誤差值。分類器一所得之分類案件數為27 件案例，錯誤率為0.0294。假設顯著水準 $\alpha=0.05$ 時，可接受的分類器之療程分類準確率為96%，則此分類器之療程分類準確率可否滿足。由下列之計算式可知檢定結果為不否定虛無假設 (H_0)，此分類器之分類準確率於97%之標準內。圖4.7為本研究分類器二誤差統計圖。

$$\hat{P} = \frac{27}{916} = 0.0294, \alpha = 0.05$$

$$H_0: P \leq 0.03, H_1: P > 0.03$$

$$Z_0 = \frac{\hat{P} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} = \frac{0.0294 - 0.03}{\sqrt{\frac{0.97 * 0.03}{916}}} = -0.1065$$

$$Z_0 < Z_{0.95} = 1.645$$

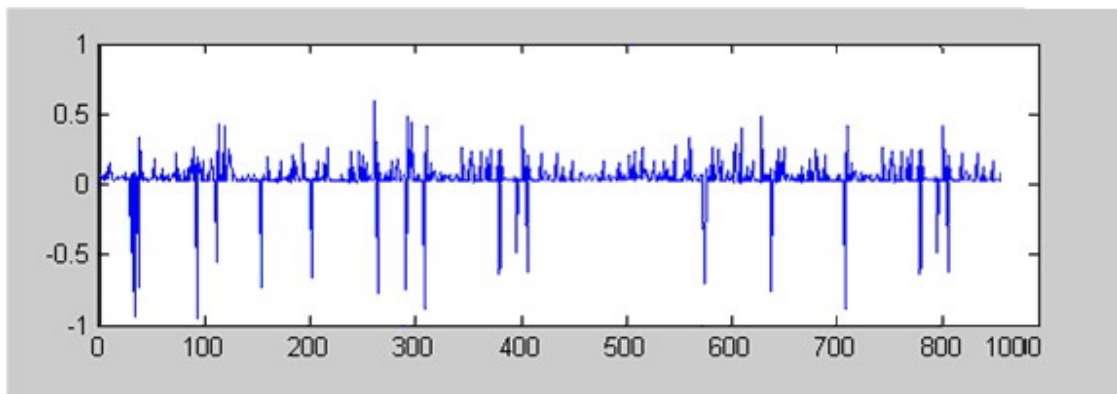


圖 4.7 分類器二誤差統計圖

由圖4.6 與圖4.7 可知，分類器二之均方誤差值範圍較分類器一之均方誤差值範圍還小，且網路模擬的分類結果準確率也高於分類器一。因此，本研究選取分類二做為心血管疾病分類器，於此網路架構下可有較佳之分類效果。

4.3 系統設計與測試

4.3.1

所使用的工具為 Microsoft Access 2003、Visual Studio.NET、ASP、Microsoft FrontPage 2003。

病歷資料庫

將 916 筆病患資料有關的輸入變數轉包含健診欄位與生活習慣共 17 項及 14 種心血管疾病。將轉換後的病歷資料匯入 Microsoft Access 2003 資料庫中，完成病歷資料庫。

知識庫

將心臟科專業醫師所提供，有關心血管疾病日常作息及飲食建議輸入 Microsoft Access 2003 中。供推理機比對病患資料後如患有心血管疾病則提出適當的建議。

知識庫是系統從病歷資料庫中的資料整理出不重複的法則，每個法則包含狀態 (Condition) 和行為 (Action) 的兩個部分，健診項目為狀態、心血管疾病類別行為。

推理機

使用 Visual Studio.NET 語法建立病歷資料比對器。比對後如符合法則，則從知識庫中抓取相關的建議傳送到輸出介面。比對之後如不符合知識庫病例法則，則將者用者輸入的數據傳送到類神經網路運算。

使用 Visual Studio.NET 建立類神經網路，如輸入的病例與知識庫中的法則不相符，將所輸入的資料經由類神經網路的加權運算，預測所可能罹患的心血管疾病。並從知識庫中抓去適合的建議。之後將所得到的病例交由專家判定與修改後存入訓練資料庫當中。類神經網路分類器定時從病歷資料庫中隨機取總病例數的三分之二作為訓練資料集，重新模擬類神經網路。

輸入介面

使用 ASP 架構 web 介面，Microsoft FrontPage 2003 製作輸入網頁，將數據填寫完畢後按下 **確定送出** 則將數據送入推理機運算。若需重新輸入則按下 **重新輸入** 所有的數據則清空。圖 4.8 中的顯示空格區域為輸入欄位。

年齡：	<input type="text" value="請選擇"/>
身高：	<input type="text"/> 公分
體重：	<input type="text"/> 公斤
臨床檢驗值：	
血壓：收縮壓	<input type="text"/> mmHg
血壓：舒張壓	<input type="text"/> mmHg
飯前血糖 (Glucose AC)	<input type="text"/> mg/dl
尿酸 (Uric acid)	<input type="text"/> mg/dl
總膽固醇 (Cholesterol)	<input type="text"/> mg/dl
三酸甘油脂 (Triglyceride)	<input type="text"/> mg/dl
高密度膽固醇 (HDL-C)	<input type="text"/> mg/dl
低密度膽固醇 (DDL-C)	<input type="text"/> mg/dl
生活習慣：	
茶：平均喝	<input type="text"/> 杯/天
煙：平均吸	<input type="text"/> 根/天
咖啡：平均喝	<input type="text"/> 杯/天
酒：平均喝	<input type="text"/> 杯/天
方糖：平均吃	<input type="text"/> 顆/天
	<input type="button" value="確定送出"/> <input type="button" value="重新輸入"/>

圖 4.8 輸入介面

輸出介面

使用 ASP 架構 web 介面，Microsoft FrontPage 2003 製作輸出網頁，將推理機運算之後，由知識庫得到的心血管疾病種類與建議顯示出來。圖 4.9 中區域一為身體質量指數，顯示是否過重。圖 4.9 中區域二為可能罹患的疾病，當推理機運算後所可能罹患的心血管疾病將會在此區域顯示。圖 4.9 中區域三為改善建議，當運算出所可能罹患的心血管疾病後，會從知識庫中抓取適合的建議在此區域顯示。

線上健康處方	
身體質量指數：	區域一
可能罹患有：	區域二
改善建議：	區域三

圖 4.9 輸出介面

4.3.2 測試

系統建立完成之後隨機選取測試資料集中的病例資料加以測試。測試資料：男性、40歲、172公分、70公斤、收縮壓150mmHg、舒張壓95 mmHg、飯前血糖98mg/dl、尿酸5 mg/dl、總膽固醇212 mg/dl、三酸甘油脂117 mg/dl、高密度膽固醇72 mg/dl、低密度膽固醇140 mg/dl、2杯茶/天、2根煙/天、咖啡0杯/天、酒0杯/天、方糖3顆/天。圖4.10為輸入測試資料系統介面。

年齡：	<input type="text" value="40"/>	
身高：	<input type="text" value="172"/>	公分
體重：	<input type="text" value="70"/>	公斤
臨床檢驗值：		
血壓：收縮壓	<input type="text" value="150"/>	mmHg
血壓：舒張壓	<input type="text" value="95"/>	mmHg
飯前血糖 (Glucose AC)	<input type="text" value="98"/>	mg/dl
尿酸 (Uric acid)	<input type="text" value="5"/>	mg/dl
總膽固醇 (Cholesterol)	<input type="text" value="212"/>	mg/dl
三酸甘油脂 (Triglyceride)	<input type="text" value="117"/>	mg/dl
高密度膽固醇 (HDL-C)	<input type="text" value="72"/>	mg/dl
低密度膽固醇 (DDL-C)	<input type="text" value="140"/>	mg/dl
生活習慣：		
茶：平均喝	<input type="text" value="2"/>	杯/天
煙：平均吸	<input type="text" value="2"/>	根/天
咖啡：平均喝	<input type="text" value="0"/>	杯/天
酒：平均喝	<input type="text" value="0"/>	杯/天
方糖：平均吃	<input type="text" value="3"/>	顆/天
	<input type="button" value="確定送出"/>	<input type="button" value="重新輸入"/>

圖 4.10 測試資料系統介面

所得結果如圖4.11所示。

身體質量指數：正常。

可能罹患：第一期高血壓。

改善建議：飲食應清淡，增加蔬菜水果的攝取。低鹽、低動物性脂肪。多喝水，適度的運動。飲食應清淡，增加蔬菜水果的攝取。低鹽、低動物性脂肪。多喝水，適度的運動。多攝取不飽和脂肪酸、大豆製品、糙米、燕麥、薏仁、儘量以魚，雞肉取代豬，牛肉。適度運動，多喝水。攝取適量的菸鹼酸、維他命C、維他命E。使用者可根據系統所提供的建議改善日常生活習慣、飲食方面，改善心血管疾病對於健康的危害，達到個人健康管理的目的。

線上健康處方

身體質量指數：體重正常。

異常指數：
體重正常。第一期高血壓。第一期高血壓。
低密度膽固醇略高。

可能罹患有：
請盡速就醫

改善建議：
飲食應清淡，增加蔬菜水果的攝取。低鹽、低動物性脂肪。多喝水，適度的運動。飲食應清淡，增加蔬菜水果的攝取。低鹽、低動物性脂肪。多喝水，適度的運動。多攝取不飽和脂肪酸、大豆製品、糙米、燕麥、薏仁、儘量以魚，雞肉取代豬，牛肉。適度運動，多喝水。攝取適量的菸鹼酸、維他命C、維他命E。

圖 4.11 測試結果介面圖

第五章 結論與建議

飲食日漸西化，罹患心血管疾病的年齡也隨之逐漸下降。國人嚴重缺乏對於心血管疾病的認知。為了讓國人能夠自行依據健康檢查報告判斷自己的健康狀況。本研究目的在於探討使用資料採礦資訊系統應用於個人健康管理上，達到管理個人健康狀況，進而提升醫療品質與降低社會成本。

近年來，人工智慧分析工具如類神經網路，已受到許多研究者之偏好，除了其不需要對資料型態進行假設驗證，且其在於捕捉資料間非線性關係的能力上，亦受到肯定。

5.1 研究成果

本研究透過分析健診資料與生活習慣項目，運用類神經網路預測可能罹患的心血管疾病並提出適當建議。採用專家知識與類神經網路建構一套心血管病徵與診斷系統。將經由專家與MIFS篩選後的特徵，利用倒傳遞網路(BPN)讀取專家知識庫資料來分析預測並且提供有關心血管疾病患者日常生活飲食相關知識。

本研究將一般國人難以涉入的專業醫學範圍轉換成簡單的介面，即使未受過專業訓練仍然可以方便使用。希望能提高國人對於心血管疾病的認識，方便日常生活習慣與飲食的管理，降低心血管疾病對於國人的危害，進而間接做為台灣地區民眾預防或延緩心臟血管疾病發生之參考。

5.2 建議

1. 影響心血管健康的因素有很多，包含生活習慣、遺傳病、運動、居家環境、用藥習慣、工作壓力、生理因素、心理因素等許多方面。由於時間、人力、成本等因素的限制，可收集更多的資料加以建立模型，提高準確率。
2. 後續研究可依倒傳遞網路演算法將此系統發展成家庭醫學專家系統，只需將其他科別的檢驗檢查的數據納入系統中，依照實證醫學的疾病準則去分析出為何種疾病。
3. 尋找更高效率的群集演算法，以提高療程群集之效率。因倒傳遞演算法收斂速度很慢，所以學習速度也很慢。其建構模式時所需花費之訓練時間相對較長，且其在變數重要程度的判讀上亦屬不易，由於類神經網路在不同研究下之最佳參數組合（如：學習速率、隱藏層層數、隱藏層神經元數目等）並無一定法則可遵循，通常是藉著試誤的方式找出，未來朝此研究，找出適用本研究領域之網路架構。



參考文獻

- 【1】中央健康保險局 <http://www.nhi.gov.tw/>
- 【2】王進德、蕭大全。(2001)。類神經網路與模糊控制理論入門。全華出版社。
- 【3】行政院衛生署資訊網 <http://www.doh.gov.tw/cht/index.aspx>
- 【4】江韶岳(2004)。心臟病健康指南。原水
- 【5】朱佳雯(2003)。案例式推理與類神經網路在心電圖診斷之應用研究。私立真理大學管理科學研究所。碩士論文。
- 【6】李淑芬(2002)。臨床路徑之建立機制—應用資料採礦技術。東海大學工業工程與經營資訊研究所碩士論文。
- 【7】林傑斌、劉明德(2002)。資料採掘與OLAP 理論與實務。文魁資訊股份公司。
- 【8】吳國禎、趙一平、蘇振隆(2002)。資料探索方法在醫學資料庫之評估。中原學報，30(1)，51-61。
- 【9】哈佛健康健診 <http://www.hvc.com.tw/>
- 【10】洪哲倫(2001)。Data-Mining 遠景 無限寬廣。網際先鋒，138-141。
- 【11】黃仁貴(2002)。以叢集分析技術探討病患就診屬性與看診時間之關係。台北醫學大學醫學資訊研究所碩士論文。
- 【12】湯玲郎、林信忠(2000)。資料萃取法在健保費用稽核之研究。醫療資訊雜誌，11，85-104。
- 【13】財團法人長庚紀念醫院全球資訊網 <http://www.cgmh.org.tw/>
- 【14】孫漢屏(2002)。類神經網路為基礎之智慧型醫院網路掛號系統。中國醫藥學院醫務管理研究所。碩士論文。
- 【15】凌群電腦股份公司(2000)，「物件導向技術與實例」，儒林圖書有限公司，。
- 【16】唐壽生(2003)。資料探勘技術應用於肺結核病患完治的預測。國立中正大學資訊管理學系。碩士論文。
- 【17】袁繼銓(2002)。以類神經網路預測燒傷病患住院日之研究。國立中山大學資訊管理研究所。碩士論文。
- 【18】姚步慎(1991)，「人工智慧與專家系統導論」，基峰資訊股份有限公司。
- 【19】莊莉瑩(2001)，「資料挖掘機制在臨床路徑之運用」，私立東海大學工業工程學系。碩士論文。
- 【20】國立台灣大學附設醫院 <http://ntuh.mc.ntu.edu.tw>

- 【21】雷賀君(2003)，前十字韌帶傷害快速診斷系統-以粗略集合、基因演算法與倒傳遞網路為工具。大葉大學工業工程學系。碩士論文。
- 【22】張維國、丁予安、陳懋良（1989）：臺灣地區動脈粥樣硬化疾病之營養及其他相關危險因素之探討。中華醫誌；44：191-7。
- 【23】楊超然(2004) 利用文件及影像檢索建立胃癌診斷與治療的案例式推理。私立台北大學醫學資訊研究所。碩士論文
- 【24】瑞輝大藥廠 <http://www.pfizer.com.tw/>
- 【25】葉怡誠（1998）。類神經網路模式應用與實做。儒林出版社。
- 【26】蔣肇慶、林熙禎（1999）。資料開採在醫療資訊的研究。醫療資訊雜誌，9，71-82。
- 【27】羅華強(2001)。類神經網路。清蔚科技。
- 【28】Aslandogan, Y. Alp, (2004) , “*Evidence combination in medical data mining* ,” International Conference on Information Technology, ITCC, v 2, , p 465-469
- 【29】Assmann, G., and H. Schulte, (1992) , “ *Relation of high-density lipoprotein and triglyceride to incidence of atherosclerotic coronary artery disease* ,” Am J Cardiol ;70:733-7.
- 【30】Battiti, R. , (1994) , “*Using Mutual Information for Selecting Features in Supervised Neural Net Learning*,” IEEE Transactions on Neural Networks, Vol.5, No.4, 537-550 。
- 【31】Blahut, R. E. , (1987) , “ *Principles and Practice of Information Theory*,” Addison Wesley 。
- 【32】Castelli ,WP , (1986) , “*The triglyceride issue: a review from Framingham*,” Am Heart J 95; 112:432-7.
- 【33】Castelli ,WP , and K. Anderson , (1986) , “ *A population at risk: prevalence of high cholesterol levels in hypertensive patients in the Framingham study*,” Am J Med ;80:S23-S32.
- 【34】Castelli ,WP , (1992) , “*Epidemiology of triglycerides: a review from Framingham*,” Am J Cardiol ; 70:3H-9H.
- 【35】Chang, Chuan-Yu, (2004) , “*A Contextual-based Hopfield Neural Network for medical image edge detection*,” IEEE International Conference on Multimedia and Expo , v 2, p 1011-1014
- 【36】Gordon ,T , WP Castelli, MC Hjortland, WB Kannel, and TR Dawber. ,(1977) , “ *High-density lipoprotein as a protective factor against coronary heart disease: the Framingham study* ,” Am J Med ; 62:707-14.
- 【37】Guler, Inan, (2005) , “*Improving medical diagnostic accuracy of ultrasound Doppler signals by combining neural network models*,” ,Computers in Biology and Medicine, v 35, n 6,

July, 2005, p 533-554

【38】 Hjermann, I, BK Velve, I Holme ,and P. Leren ,(1981) , “*Effect of diet and smoking intervention on the incidence of coronary heart disease,*” Report from the Oslo Study Group of a randomized trial in healthy men. *Lancet* ; 2:1303-13.

【39】 Ichimura, Takumi, (2003) , “*A classification capability of Reflective Neural Networks in medical database,*” *Lecture Notes in Artificial Intelligence*, v 2774 PART 2, p 373-379

【40】 Kagan ,A, (1974) , “*Epidemiologic studies of coronary heart disease and strokes in Japanese men living in Japan ,*”Hawaii and California: demographic, physical, dietary and biochemical characteristics. *J Chron Dis* ; 27:345-64.

【41】 Katz , J.O. ,(1992) , “*loping nearul network forecaster for trading,*”*Technical Analysis of Stocks and Commodities* , p.58~70.

【42】 Keys, A, A Menotti, C Aravanis, H Blackburn, BS Djordevic, R Buzina, AS Dontas, F Fidanza, MJ Karvonen,and N Kimura,(1984) , “*The seven countries study: 2,289 deaths in 15 years,*” *Prev Med* ; 13:141-54.

【43】 Lehmann, Thomas M . , (2005) , “*Automatic categorization of medical images for content-based retrieval and data mining,*” *Computerized Medical Imaging and Graphics*, v 29, n 2-3, March/April, *Imaging Informatics*, p 143-155

【44】 Leong, Jenn-Hwai ,and Mei-Ming Kuan, (2005) , “*A hybrid neural network system for pattern classification tasks with missing features Lim,*” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v 27, n 4, April, 2005, p 648-653

【 45 】 Megaputer , (2002) , “*Medical Fraud Detection Through Data Mining,*” www.megaputer.com

【46】Pyle Dorian ,(1999) , “*Data Preparation for Data Mining,*” Morgan Kaufmann, California.

【47】 Shaout, Adnan, (2003) , “*A fuzzy data mining system for medical records,*” *Proceedings of the International Conference on Artificial Intelligence IC-AI 2003*, v 1, 2003, p 376-381

【48】 Villiers, J. , and E.Barnard, (1993) , “*Back-Propagation neural nets with one and two Hidden Layers,*” *IEEE Transaction On Neural Networks*, Vol. 4, No. 1, pp. 136-141.

【49】 Yano K, DM Reed , and DL McGee , (1984), “*Ten-year incidence of 100 coronary heart disease in the Honolulu Heart Program: Relationship to nutrient intake ,*” *Am J Epidemiol* ; 119:667-76.

心血管性疾病及營養攝取

冠狀動脈心臟病(Coronary Heart Disease)

一、危險因子：

包括年齡、性別、遺傳、抽煙、高血壓、肥胖症、高脂血症、高血糖與糖尿病、運動、體型、感情壓力及個性。

二、血脂過高與冠狀動脈心病：

高血脂症是血液中的膽固醇、三酸甘油酯增加，一般高膽固醇血症是指血清膽固醇濃度超過 240 毫克/100 毫升。

三、影響血脂肪的飲食因素：

膽固醇、脂肪的類型、纖維、鈣、咖啡、酒精、減重。

高脂血症之飲食治療原則：

1. 控制熱量的攝取，以達到理想體重。
2. 限制膽固醇的攝取量。
3. 少吃高飽和脂肪的食物。
4. 減少油脂之攝取量。
5. 多吃高纖維的食物。
6. 少吃甜食並限制酒量。

脂肪酸油脂的組成分子，依照其飽和度為：飽和、單元不飽和及多元不飽和脂肪酸三類。

飽和脂肪酸多存在於動物性油脂中，且在室溫下常呈固態。
每日總攝取量應低於 300 毫克。

膽固醇只在動物性食品中才有，植物性食品中不含膽固醇。
各種肉類（包括：雞、鴨、魚、豬、牛、羊等）
膽固醇含量：平均每兩約含 20~30 毫克膽固醇。
若每日攝取肉類 5 兩左右，則膽固醇攝量在 200 毫克以下。

飽和脂肪酸多半存在下列食物中：

動物性食品：肥肉、全脂奶、奶油、豬油、牛油、肉類外皮

植物性食品：烤酥油、椰子油、椰子、棕櫚油。

除少吃上述食物外，烹調用由宜選擇較不飽和的油。

例如：大豆、米、油、玉米油、紅花籽油、葵花仔油、蔬菜油、橄欖油、花生油、芥花油、苦茶油。另外，魚類及豆類之飽和脂肪含量較少，亦可多考慮用以取代其他肉類，作為蛋白質之來源。

不吃或盡量少吃：

高油食物：肥肉、各種肉類之外皮、香腸、豬腸、牛腩、培根、油炸肉塊、油條、蛋餅、起司、奶油、肉湯、沙拉醬。

高油點心：腰果、花生、瓜子、蛋糕、西點、中式糕餅、巧克力、冰淇淋除上述食物應限制外，烹調用的植物油應適量使用，儘量採取低油之烹法：每餐只吃1~1.5匙油，1湯匙=15公克

高血壓病

高血壓是個無形的殺手，在平常它不會產生任何的症狀，當高血壓發生非常嚴重的時候，它可以產生頭痛、頭暈、以及流鼻血。高血壓是造成心臟病死亡、腦中風死亡、以及尿毒症的一個主要的原因。

食物種類	可食的食物	應避免的食物
麵包、穀類、麵油類及脂肪	脫脂牛奶、茶、咖啡、雞蛋、汽水皆適可。植物油：花生、玉蜀黍、大豆等	全脂牛奶、全脂牛奶動物脂肪, 牛、豬、羊脂牛油豬油、酥油等
肉類和家禽	瘦肉	肥肉、火腿、內臟：腦、肝、腎
魚類	無油脂類或魚肉	油脂類：魚卵、蝦、螃蟹、龍蝦
奶製品	蛋白、低脂乳酪、脫脂奶粉	蛋黃、全脂乳酪、奶油、

		牛油、酥油
湯類	蔬菜清湯	含油脂的濃湯
蛋糕甜點	蛋白甜餅、餅乾、蜜糖	布丁、豬油或蛋黃製成的點心、雪糕
水果及蔬菜	可以	無須避免

高血壓的飲食

- 一、血壓和食鹽有直接的關係。影響血壓其實是存在鹽中的鈉質。患有高血壓的人，一日食鹽的份量是不可以超過7克。
- 二、有研究指出，有家族性高血壓的人，增加鉀質的攝取能有效降低血壓。含豐富鉀質的食物包括蔬菜和水果。不過腎臟不好的患者可能要限制鉀質攝取量，應詢問營養師或醫生的意見。
- 三、研究指出，正常人的血壓的鈣質攝取量比患有高血壓的人為高。在一項研究中，正常人血壓若每日進食1克的鈣質，上壓和下壓都會明顯地降低。但在現階段，還未有足夠資料去建議患有高血壓的人進食多少鈣質。
- 四、除了以上因素外，肥胖和高血壓是有很大的關係，所以體重超重的患有高血壓的人，都需要減肥。

心臟血管疾病

血膽固醇在決定一個人得到冠狀動脈心血管疾病的危險性，扮演著非常重要的角色，你的血膽固醇值愈高，得到冠動脈心血管疾病的危險就愈高，這就是為什麼高膽固醇血症被稱為心血管疾病危險因子的原因。

心臟病的成因

心臟病的出現，是由於冠狀動脈內壁上的膽固醇塊上，形成了血塊，而令到供應給心臟的血液嚴重減少，引致心絞痛、心肌組織產生電脈性不穩。倘若供應給心臟某部份血液完全停止時，該部份便無法得到氧氣和賴以保持正常運作的養份，造成心肌梗塞，甚至導致患者死亡。

健康心臟的保養指引

不論你的血膽固醇值如何，你都可以改變血膽固醇值或將它維持在理想的指數來降低得到心臟血管疾病的危險性。下列是維護健康心臟的幾個必要條件，適用於全家人：

(1) 食用低飽脂肪酸以及低脂肪的食物

食物中含有脂肪，而這些脂肪是由飽和脂肪酸加上不飽和脂肪酸所混合而成。飽和脂肪酸會提高血膽固醇。降低血膽固醇最好的方法就是選擇低飽和脂肪酸的食物。許多全脂的食物，大部份含有很高的飽和脂肪酸，所以選擇低脂肪的食物有助於家人吸收較少的飽和脂肪酸。當你食用油脂時，可以用未飽和脂肪酸來代替飽和脂肪酸—不論是多元不飽和脂肪酸或單元不飽和脂肪酸都可以。脂肪是熱量（卡路里）的豐富來源，所以吃低脂肪的食物會幫助你吃進較少的熱量。如果你的體重超重的話，吃較少熱量可以幫助減輕你的體重。

(2) 食用高澱粉，高纖維的食物

含有高澱粉及高纖維的食物是極佳的飽和脂肪酸食物之替代品。像是麵包、穀物類、通心麵、水果及蔬菜等都是低飽和脂肪酸、低膽固醇之食物。同時，它們所含之熱量也比高脂肪的食物少了許多。像奶油及沙拉的醬料及全脂牛奶都含有高熱量及高脂肪量。高澱粉及高纖維之食物是礦物及維他命的主要來源。食用含有可溶性纖維素的食物像燕麥、麩糠類及乾豌豆、蠶豆等這些食物，可以幫助你降低血脂肪。

(3) 食用低膽固醇的食物

食用含有膽固醇的食物，會使血中膽固醇升高。飲食中的膽固醇主要來自肉類食物。一些肉類食物飽和脂肪酸含量雖低，但有可能含有高膽固醇，像肝臟及蛋黃這類的食物，都是屬於低飽和脂肪酸，但是卻含有高量的膽固醇。蛋白及植物性的食物，則不含膽固醇。

(4) 注意豆類製品的攝取

臨床研究顯示，飲食中以大豆蛋白取代動物性蛋白，可以降低血膽固醇（特別是低密度脂蛋白膽固醇）、三酸甘油酯等，因為大豆蛋白中含有異黃酮素、皂素，可以減少血脂肪的氧化，並降低膽固醇的吸收。每天攝取 25—50 克的大豆蛋白，可以相對降低 8.9—17.5mg/dl 的血膽固醇。

(5) 適度的運動

運動可以幫助改善血膽固醇之指數：它可以增加“好的膽固醇”及降低“壞的膽固醇”。多做運動也可以幫助減輕體重、降低血壓、適度改善心臟血管的健康，並能減輕壓力。全家一起運動，也是很好的方法。

(6) 保持理想的體重—如果體重超重，就要減重

體重過重的人比標準體重的人容易傾向血中膽固醇值較高，體重超重又具備“蘋果身材”的人比“梨型身材”的人得到心臟病的危險性高。不論你的身材的何種型，減少食用油脂，就可減少熱量，因為油脂是熱量最豐富的來源。一個家庭之飲食習慣若改以高澱粉及高纖維的食物替代高脂肪食物，是一個控制體重的方法。

中 風

中風是腦血管疾病的俗稱。腦梗塞的原因相當複雜它可以分為：血栓(thrombosis)，栓塞(embolism)，及小洞梗塞(lacunar infarct)等。

◆ 導致中風的因素

- (一) 腦栓塞—腦血管被血栓所阻塞
- (二) 腦血管梗塞

◆ 中風患者飲食所要注意的事項：

1. 多吃高纖維質食物
2. 避免進食高飽和脂肪的食物
3. 減少進食高膽固醇的食物
4. 避免過鹹或過甜的食物
5. 適當烹調方法能降低
6. 戒菸、戒酒

◆ 如果有中風的前兆，除了要找醫師就診之外，日常飲食上要注意以低鹽、低脂肪、低膽固醇為主；禁食辛辣刺激食物；戒煙、酒，另外可以多補充以下食物：

1. 可以多攝取豐富蛋白質：建議：可以適量選用雞蛋、脫脂牛奶、魚、瘦肉及豆類製品等善加補充。
2. 要多吃高纖食品：建議：富含纖維素的蔬果，包括有豆芽、芹菜、韭菜、菠菜、大白菜、空心菜、黃瓜、冬瓜、番茄、蘿蔔、海帶等。
3. 要多吃降血壓食物：建議：一般的蔬菜、水果
4. 多吃降脂食品：建議：黃豆類製品、優酪乳、蜂王漿、洋菇、苜蓿、海藻...等都是相當好的補充來源。

◆ 減少攝取

1. 減少油脂攝取：在選擇動物性食物時，要記得這個優先順序：魚肉 > 禽(雞)肉 > 豬、牛肉

2. 少吃鹽：當我們吃麵時，麵湯當中所含的鹽份也很高，因此最好遵守「吃麵留湯」。

肥胖

定義：聯合國衛生組織曾對肥胖下定義：當一個人體重超過理想體重得20%時，“肥胖”；超過10~20%者，稱為“體重過重”。運動是保持人體健康的重要一環、也是苗條過程中重要的一部分。肥胖應該與熱量攝取過多、人對食物熱量利用差，缺乏足夠運動來消耗熱量，以致儲存於體內而造成肥胖。以運動配合飲食控制，可以提升心肺功能、減少減重的不適、更可改善基礎代謝率的降低，而達到持續苗條效果。

無論男女，都可以BMI來判斷肥胖的程度公式：

BMI=體重/身高(公尺)的平方

BMI=20~24.9 0級(正常)

BMI=25~29.9 I級(輕度肥胖)

BMI=30~40 II級(中度肥胖)

BMI>40 III級(重度肥胖)

肥胖的類別：“肥胖”主要是由於體內脂肪組織增加的結果；脂肪組織的增加可以是脂肪細胞內油脂量不斷增多的「細胞肥大」或是「細胞數目增多」，或是上述兩種狀況合併發生的結果。人的一生當中，有幾個階段事屬於細胞快速生長的時期，如：嬰兒期、幼兒期、青春發育期、懷孕期。在這幾個階段，吃得過多容易造成脂肪細胞的大量增加。女性化的肥胖是指多餘的脂肪主要積存在臀部和大腿上，而男性化肥胖則在胸部、腹部和手臂上。

(一) 烹調注意事項

1. 儘量避免使用精緻糖，必要時使用「代糖」調味
2. 以無油烹調為原則，如：清蒸、水煮、烤、清燉、滷、涼拌等
3. 避免勾芡及使用大量含糖調味料
4. 調味不宜太鹹

(二) 可隨意食用之食物

1. 飲料：可飲用不加糖的咖啡、清茶，或低熱量之汽水和可樂。
2. 調味料：依菜式可添加蔥、薑、蒜、醬油、醋、芥茉、五香、花椒、八角、胡椒粉、肉桂、咖哩粉、檸檬等，增加美味。
3. 代糖：代糖製作糖果，可滿足對甜味的慾望。
4. 其它：無糖或代糖製作的愛玉、仙草、洋菜。

(三) 烹調原則

1. 多選用清蒸、水煮、烤、滷、涼拌、清燉等方法烹調。避免油炸、油煎、油炒等，以減少油脂的攝取。
2. 蔬菜：以去油高湯、水燙等方法烹調，攝取量依照醫師指示。各種蔬菜配製的生菜沙拉如：小黃瓜、蕃茄、蘆筍、青花菜、白花菜、菠菜、四季豆、生菜葉、白的或紫的包心菜等都是很好的選擇。
3. 烹調時，應避免使用糖醋、蜜汁、醋溜、油炸等方式。
4. 烹調時宜清淡，避免攝取過鹹的食物。