智慧型為基之糖尿病線上檢測及風險評估輔助系統建立之研究

The Study of Constructing an Intelligence–Based On-Line Support System for Diabetes Diagnosis and Risk Evaluation

邱瑞科 [a*], 黃珈瑩 [a], 涂秀美 [b], 張彥群 [c]
Ruey-Kei Chiu [a], Jia-Ying Huang [a], Xiu-Mei Tu [b], Yan-Qun Zhang [c]
[a] 輔仁大學資管所, [b] 恩主公醫院社區醫學部, [c] 恩主公醫院資訊課
[*]通訊作者： 邱瑞科, rkchiu@mail.fju.edu.tw

摘要

糖尿病發生率逐年增加，近來已備受關注與研究。糖尿病初期往往會因不易察覺而被忽略，一旦罹患便須長期治療控制，且現今大部份的身體健康檢查，並未能真正達到風險評估及預防保健的真正效益，若能藉由一診療輔助系統的開發建構，則可協助醫學臨床之早期診斷與評估。

本研究藉由文獻探討及彙整醫師專家的專業知識，以釐清引發糖尿病之顯著危險因子，並嘗試運用人工智慧技術中之類神經模糊系統模型與模糊專家系統之整合應用，來發展一可用於線上檢測及風險評估輔助系統，以提供醫師作為糖尿病風險評估、了解患者病情之資訊給予適當生活調適指引及更精準診療之參考依據。

本研究所建立的糖尿病檢測及決策輔助系統，經實驗結果顯示，類神經模糊系統診斷其分類正確率達88.41%，顯示系統可以達到與專業判斷的水準，藉由此一評估系統可透過預防醫學的角度，深入分析每個人最切身的健康風險，提供早期治療的契機。

關鍵字：人工智慧、類神經模糊系統、模糊專家系統、糖尿病。

Abstract

The incidence of diabetes has increased every year and recently it has been concerned and frequently studied. It is difficult to detect the disease early whiles most of the physical examination don't achieve the effectiveness of risk assessment and preventive health care. Once suffering the disease, people must be controlled for long-term treating. Therefore, if a diagnostic support system may be built, it can early assist in clinical diagnosis and detection.

By reviewing literature and collecting professional knowledge of the physician, it will clarify risk factors of causing diabetes. This study attempts to apply the integration of the neural fuzzy system model and fuzzy expert system in artificial intelligence technology to develop an online detection and risk evaluation support system for detecting the serious degree, and provide to risk evaluation for diabetes. By means of this system, a physician may give a more precisely diagnosis, whiles the right guidelines of health education and lifestyle adjustment to the diabetes may be given.

The results of this study show that decision support system for diabetes diagnosis and risk evaluation can achieve 88.41% of classification and it can be judged with the professional standards, analyzing personal health risk deeply through the perspective of preventive medicine.

keyword：Artificial Intelligence, Neural Fuzzy System, Fuzzy Expert System, Diabetes.

1. 前言

糖尿病是一種慢性的代謝異常疾病，近年來由於老年化的社會結構，加上飲食與生活習慣的改變，不僅導致糖尿病發病率逐年增加的現象，在台灣更數年位居我國十大死因第四位，因此相關之疾病之育早期發現與預防已刻不容緩。

由於糖尿病容易造成許多併發症的產生，除了在醫療照護應努力改善糖尿病的控制，減少併發症的發生外，也應加強預防糖尿病的發生或儘早篩檢出疾病輕微的潛伏糖尿病病患，以避免疾病惡化後，造成民眾生活品質的降低及醫療支出的沈重負擔[1]。

然傳統臨床判定糖尿病雖有一套診斷數值，有時檢驗同一個人身上，卻時常出現不一致的現象[3]，而專業醫護人員以人力判讀長時間累積的數據，也可能產生因人而異的判斷標準；有鑑於此本研究希望針對糖尿病患作為研究主題，利用人工智慧技術之類神經模糊系統檢測其罹患之分類結果，再藉模糊專家系統加以處理與整合，發展一套糖尿病檢測及決策輔助系統，運用模糊推論方法進行分析，提供正確而有效的糖尿病風險評估，提升個人瞭解自身健康狀態的風險，以降低罹患疾病機率或者助於盡早就醫的情形。

將藉由資訊系統及網路的便利性，能隨時隨地進行風險評估與分析，並且得到適當的衛教資訊，研究對象不只侷限在已罹患糖尿病的病患，協助減少併發症的產生或惡化，對於潛在的高危險族群與一般健康民眾而言，在初期預防階段可瞭解罹患糖尿病的風險等級，以及對引發疾病之危險因子有正確的認知並有效地控制，以便有效評估和預防，輔助給予適當處置如正確生活型態衛教與調適，甚或積極的治療，以達到糖尿病積極預防之成效。

2. 研究方法

圖 1 描繪本研究所擬發展之糖尿病線上檢測及決策輔助系統架構圖。

2.1 糖尿病危險因子之確定及分析

主要為提供本研究制定模糊法則之依據與相關背景知識的來源，藉由探討文獻資料，以及訪談專業醫護人員所得到確切肯定的醫療知識，將影響糖尿病之危險因子加以彙整後，利用層級分析法(Analytical Hierarchy Process, AHP)透過 Expert Choice2000 軟體進行檢驗及分析，以計算出糖尿病影響危險因子之權重，歸納整理推論系統的初步建置作業。因子權重值整理如表 1、表 2 所示。

表 1 各判定因子佔整體之權重值

整體排名	判定因子名稱	整體因子權重比
1	飯後兩小時血糖	0.073
2	糖化血色素	0.070
3	飯前血糖	0.058
4	隨機血糖	0.024

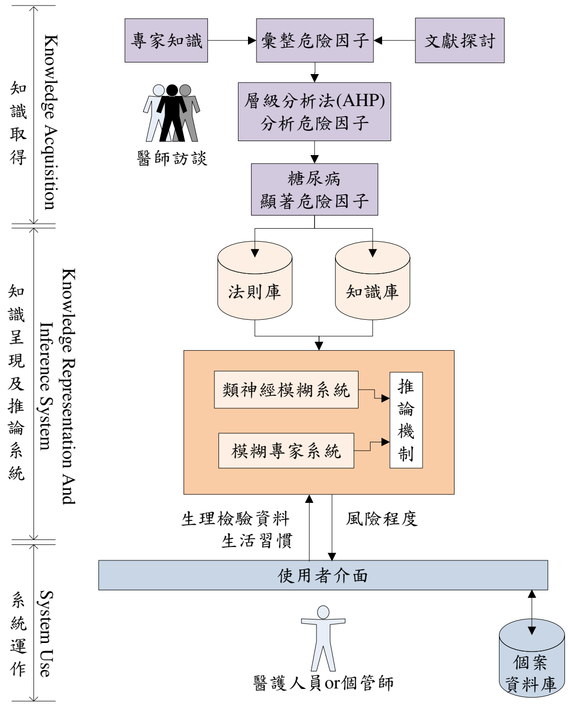

圖1 糖尿病決策及檢測輔助系統架構圖

表2 各影響因子佔整體之權重值

整體排名	影響因子名稱	整體因子權重比
1	飲食	0.106
2	運動	0.094
3	吸煙	0.075
4	腰圍	0.071
5	糖尿病家族史	0.058
6	高密度脂蛋白	0.045
7	身體質量指數	0.044
8	三酸甘油脂	0.044
9	血壓	0.042

10	年齡	0.037
11	喝酒	0.037
12	經常熬夜	0.036
13	總膽固醇	0.029
14	尿蛋白	0.028
15	高血壓家族史	0.019
16	尿糖	0.013

2.2 模糊推論

本研究為使用類神經模糊系統(Neural-Fuzzy System)進行糖尿病之風險評估部分，此方法乃目前應用於類神經網路架構上發展模糊理論，能有效的去學習並提供模糊規則(If-Then Rule)，將模糊邏輯的推論與類神經網路的學習加以整合。類神經網路提供演算法作為低階層級的學習、分類與最佳化；模糊邏輯則將推理置於語意與語言的高階層級，所以兩者是理想的互補性結合，預期可以將類神經網路低階的運算與學習能力帶引至模糊邏輯的高階層級[4]。

由於類神經模糊系統結合兩者技術上的優點，它建立在具備認知層面的模糊規則庫，而且經由學習能夠調整權重，藉著支持度的調適來消除模糊規則庫與實際運作時產生的誤差[2]。透過此技術不僅僅只能歸納，還可以進一步解釋人類依照經驗法則所產生範例的輸入輸出值之間的因果關係，因此，與其它的類神經技術相比較，類神經模糊系統有下列之優點為1. 是從模糊邏輯的預先建構的系統開始，因此學習的自由度受到控制，避免錯誤學習。2. 繼承自模糊邏輯系統的先驗知識（Prior Knowledge），因此可以解讀或加以推論學習所得到的結果[5]。

其架構如圖2所示，本研究將依據界定糖尿病的診斷檢驗項目所建立法則知識庫，並訂定危險因子的模糊區間進而導出其歸屬函數，藉由使用者輸入變數資料後，透過模糊推論執行法則比對及觸發動作，推論出最後的圖形結果以計算其輸出結果，每一層建置步驟將於後詳細敘述之。

3. 系統建置

本研究為使用人工智慧技術中之類神經模糊系統並整合模糊專家系統來進行線上決策及檢測糖尿病的推論運行，並依步驟分別加以說明。

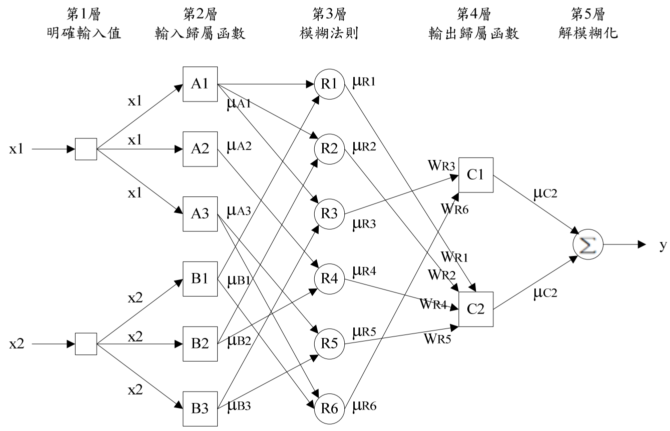

圖 2 類神經模糊系統架構[4]

3.1 糖尿病輔助檢測

此為初步篩檢罹患糖尿病之有無，使用NeuroSolutions模擬軟體建構類神經模糊網路，以進行網路訓練與測試，找出最佳模型，再應用至後續研究的模糊專家系統。

依據美國糖尿病學會(American Diabetes Association, ADA)於 1997 年公布的診斷標準，若符合下述三項中的任一項，並經確定即可診斷為糖尿病。

1. 有糖尿病症狀加上任意血糖值等於或高於 200 mg/dl。
2. 空腹血糖值等於或高於 126 mg/dl。
3. 口服葡萄糖耐受性試驗的 2 小時血糖值等於或高於 200 mg/dl。

因此，依照上述資料並經由醫療專家的認可，將空腹血糖、葡萄糖耐量試驗等設為此階段的類神經網路之輸入變數，而輸出變數為一項即罹患糖尿病之有無結果，其它的危險因子則為後續作模糊推論時所增加的輸入變數。由糖尿病之判定因子作為輸入變數，樣本資料先經正規化的處理動作，便依據各模型分別進行實驗，挑選實驗當中的最佳網路參數組合，最後評估其績效結果以驗證最佳網路模型。

3.2 糖尿病風險評估

此為藉由模糊專家系統以進行風險評估的階段，依據步驟 1.模糊化(輸入明確值與歸屬函數)、2.定義模糊規則、3.模糊推論(輸出歸屬函數)、4.解模糊化等整合成多層的網路結構[4]。

(1) 模糊化

由於本研究在系統進行檢測疾病罹患之分類結果時，

已將空腹血糖、糖化血色素等糖尿病判定因子作為模擬實驗的輸入變數，因此模糊專家系統僅利用其影響因子來評估這些因子對於糖尿病影響的風險程度。

然而研究的樣本資料為一般診療記錄，對於生活習慣或家族疾病史皆未有確切的資料數據記載之，因此，為了進一步探討其生活習慣的改善是否可影響糖尿病罹患之風險程度，故本研究將進行兩階段的風險評估設計，第一階段為個案的一般生理檢驗資料，第二階段則為假設性模擬其個案的生活習慣資料，以此比較兩階段所得的風險值結果，評估生活習慣的好壞是否能影響其罹患風險程度。

研究依權重值高低所選取變數，第一階段採用表 2 的腰圍、高密度脂蛋白、三酸甘油脂、血壓與年齡共五項輸入變數，身體質量指數因與腰圍皆同為衡量肥胖的準則，故未選取該項；而第二階段則為本研究模擬評估，將影響糖尿病所佔甚高比率的前三項排名「生活習慣」，包含飲食、運動、吸煙三者並加入第一階段的輸出風險程度，作為本階段的四項輸入變數。各項變數及其範圍的制定為經由專業之領域專家所確認驗證，如表 3 所示。

表 3 語言變數及其範圍

語言變數		語言值	模糊範圍
年齡		幼年(child)	<19 歲
		青年(young)	18-30 歲
		壯年(summer)	25-45 歲
		中年(middle)	40-65 歲
		老年(old)	>60 歲
腰圍	男性	正常(normal)	<90 公分
		稍粗(thick)	>85-95 公分
		過粗(over thick)	>93 公分
	女性	正常(normal)	<80 公分
		稍粗(thick)	>75-85 公分
		過粗(over thick)	>83 公分
血壓(收縮壓)		低(low)	<100 mmHg
		正常(normal)	95-135 mmHg
		高(high)	130-155 mmHg
		非常高(very high)	>145 mmHg
高密度脂蛋白	男性	正常(normal)	>35 mg/dl
		過低(over low)	<40 mg/dl
	女性	正常(normal)	>45 mg/dl
		過低(over low)	<50 mg/dl
三酸甘油酯		正常(normal)	<150 mg/dl
		高(high)	145-255mg/dl
		非常高(very high)	>250mg/dl
飲食(平均每天攝取蔬果份數)		良好(good)	>4 份
		正常(normal)	3-5 份
		少(less)	1-3.5 份
		非常少(very less)	<1.5 份
運動(平均每週運動天數)		良好(good)	>3.5 天
		正常(normal)	2-4 天
		少(less)	0.5-2.5 天
		非常少(very less)	<1 天
吸煙(平均每天吸煙根數)		正常(normal)	0 根
		稍高(little high)	1-12 根
		高(high)	10-20 根
		非常高(very high)	>18 根
風險程度		非常低(very low)	<14.5%
		低(low)	14-28.5%
		普通(common)	28-42.5%
		稍高(little high)	42-56.5%
		高(high)	56-71%
		非常高(very high)	70.5-86%
		嚴重(serious)	>85.5%

以表 3 之危險因子定義歸屬函數，由於傳統集合以主觀意識判定，無法明確解釋其定義的程度差異，本研究將運用梯型歸屬函數來表示模糊化過程，例如以輸入變數「年齡」來對照兩者之不同概念(見圖 3、4)，小於 19 歲屬於幼年、18 到 30 歲屬於青年、25 到 45 歲屬於壯年、40 到 65 歲屬於中年，而超過 60 歲則屬於老年。其它變數亦依此相同方式來設定。

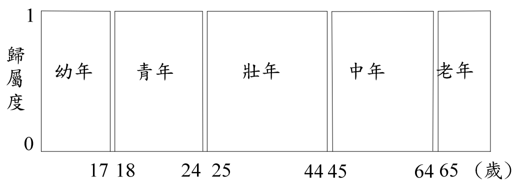

圖 3 傳統集合對"年齡"的定義

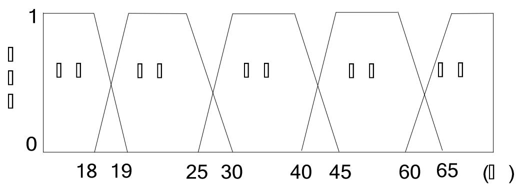

圖 4 模糊集合對”年齡”的定義

(2) 建立模糊規則

本研究藉由影響糖尿病嚴重程度的危險因子，包括年齡、腰圍、血壓、高密度脂蛋白膽固醇、三酸甘油脂及生活型態(飲食、運動和吸煙)等，這些影響罹患風險的項目建立系統之法則(見表 4、表 5)，其權重為使用層級分析法已訂定出每一因子的權重值，依其權重加總以計算出風險值的比例。經由下述模糊推論機制與解模糊的步驟之後，所得到一個明確的百分比數即表示為罹患糖尿病風險程度的機率。

表 4 第二階段之模糊規則

法則	輸入				輸出
	Step1. 風險	飲食	運動	抽煙	**Step2. 風險**
法則 1	非常低	非常少	非常少	正常	高
法則 2	非常低	非常少	非常少	稍高	非常高
法則 3	非常低	非常少	非常少	高	嚴重
…	…	…	…	…	…
法則 448	嚴重	良好	良好	非常高	普通

(3) 模糊推論

本研究使用 Java 語言的 Fuzzy Toolkit 模組建立系統，予以處理模糊邏輯規則和使用者圖形介面的系統架構。而模糊推論與解模糊的部分是採用 Mamdani 的 Min-Min-Max 推論方式[4]進行推論，藉由觸發法則後，產生相對應的圖形方式。

(4) 解模糊化

將模糊推論後之圖形以重心法之計算方法，針對所產生之圖形進行解模糊動作後，系統將輸出一個明確數值，該數值即為最後所得之罹患風險百分比率，如圖 5 所呈現之結果。

4. 實驗及結果分析

本研究蒐集北區某地區醫院之門診健檢與糖尿病個案的檢驗數據，共 638 筆樣本資料，其中 500 筆作為訓練樣本並隨機以 10%為交互驗證樣本，其餘為測試樣本以此建構類神經模糊系統，進行糖尿病檢測之診斷。

在糖尿病檢測方面，在最佳網路參數設定的組合下進行實驗，雖網路訓練的學習分類未達到百分之百，有 92.86%，而測試結果的準確率可達 88.41%、敏感度 86.96%、特異度 89.13%。

而在糖尿病風險評估方面，系統將進行兩階段的風險評估，如圖 5 的左邊部分顯示為第一階段，輸入某一個案(編號為 161379)的五項變數資料，該個案實際的疾病狀態為"有罹患糖尿病與代謝症候群"，經系統計算其嚴重程度為 66.58%屬於風險高之族群；而圖 5 的右邊部分為第二階段，因樣本資料的關係，實驗將以模擬狀況下輸出其生活習慣進行第二階段的風險評估，其中予以採計第一階段的 66.58%作共同的風險衡量，假設該個案每日攝取的蔬果僅一份、抽菸根數約有 25 根、每週實行運動僅一天，那麼若在未採取改善計畫的情況下，對於日常不正規的生活習慣而言，他的風險程度將提高到 78.25%有待改變以養成良好的生活習慣。

圖 5 糖尿病風險評估之執行畫面

5. 結論及建議

本研究透過預防醫學的角度，深入分析每個人最切身的健康風險，所發展此一檢測與風險評估系統，提供早期治療的契機，更可適時得到適當的醫療保健服務，以達到及早發現及早治療的目的，並透過生活習慣的改變與運動，來預防疾病的發生與減少發病的惡化與嚴重程度。

經過系統實證與專家評估後，其類神經模糊系統可達88.41%的分類準確率，而敏感度86.96%、特異度89.13%，更進一步說明結合類神經模糊系統與模糊專家系統的架構可適於應用在醫療疾病診斷上，是開發系統的好方法。應用本系統在糖尿病分析的領域不僅可改進醫療品質，且能達到風險評估及預防保健的真正效益，並可協助醫生對病患進行有效的臨床診斷。

而研究主要是針對一般醫療檢驗的原始資料，經過層級分析法的權重分析後，找出診斷糖尿病最具貢獻度的判定因子與影響因子，再依資料項目的特性建立類神經網路模型來模擬分析疾病的預測分類，最後透過模糊專家系統的模糊推論過程，經實證所得罹患糖尿病之風險嚴重程度。

除此，國內對於糖尿病危險因子之定義尚缺乏這類相關的調查研究，而本研究針對國內的生理及生活特質所引發的致病危險因子，經醫師專家的驗證和經驗判讀下更具有其可信度，並將這些相關因子導入於系統的推論裡使其更具完整性，因此可提供未來糖尿病或前期糖尿病發展研究的參考以釐清界定之指標。

本研究為運用類神經模糊推論系統的架構，對於模糊推論過程，其糖尿病檢測分類率與罹患風險程度，雖有不錯的預測可能發生率，亦期望日後研究的方向可朝下述幾點建議以持續發展。

(1) 本研究檢測糖尿病之影響因子時，因樣本資料的關係缺乏家族疾病史與生活習慣兩項資料，僅在風險評估階段納入生活習慣作模擬驗證，往後可利用模糊理論之概念，將此兩項資料考慮列入檢測分析時所用。

(2) 研究之實驗過程為依循Sugeno模糊模式來定義糖尿病因子的推論模糊規則，且軟體工具受限於五筆的輸入變數，可能無法全面性探討其影響關係，因此，未來可應用其他開發工具或者挑選重要特徵變數，以避免特徵維度增加而造成系統的龐大負荷，仍能精準的制定模糊規則以及充分呈現糖尿病的特徵因子，則是後續研究者可以思考的方向。

(3) 本研究主要以類神經模糊系統做為訓練模型，可利用其他類神經網路模型與之比較，找出最適當之網路模型。

致謝

本研究承蒙9 恩主公醫院資訊專家、社區醫學部之全體醫師及個案管理師之不餘遺力指導及提供研究環境及實驗樣本，研究方得以順利完成，成果亦得以彰顯，謹此致上最高謝意。

參考文獻

1. 潘文涵、傅茂祖，行政院衛生署國民健康局，取自 http://www.bhp.doh.gov.tw/health91/1-3-11-1.pdf，2008 年 8 月。
2. 蘇木春、張孝德，機器學習：類神經網路、模糊系統及基因演算法則，三版，全華科技，2004 年，頁 7-1~9-4。
3. Kahn, R., The Exert Committee on the Diagnosis and Classification of Diabetes Mellitus: Follow- Up Report of the Expert on the Diagnosis Diabetes Mellitus. Diabetes Care, 26 (11), 2003.
4. Negnevitsky, M., *Artificial Intelligence: A Guide to Intelligent Systems,* 2nd Editions. Addison Wesley, 2005.
5. PAL, S.L. and Shiu, S.C.K., Foundations of Soft Case-Based Reasoning, John Wiley & Sons, Ltd, 2004, pp. 231-252.